Integrated Project on Interaction and Presence in Urban Environments

FP6-2004-IST-4-27571

ipcity.eu

Report on Environmental Awareness application evaluation

For use as public resource

Doc-Id:	**I7.7 (adapted for public use)**
Version:	**1.0**
Author(s):	**Ann Morrison, Peter Peltonen, Saija Lemmelä, Dieter Schmalstieg, Antti Oulasvirta, Alessandro Mulloni, Antti Juustila, Holger Regenbrecht, Giulio Jacucci**
Date:	**2010-04-01**
Status:	**Final**
Availability:	**Confidential report adapted for public distribution**
Distribution:	**General Public Distribution**

Table of Content

Abstract

This internal report describes the evaluation methods used in the Environmental Awareness showcase and reports the evaluation of the MapLens prototype developed in this showcase

Intended Audience

The intended Audience of this report is IPCity internal. However, the main findings of the study will be made available to a larger audience as part of the public deliverable D7.4, and as a resource in and of itself.

1 Introduction

In this document we describe the field analysis processes used in work package 7, in particular MapLens field trials in 2009.

To gain more understanding of presence in Mixed Reality interaction, our approach has been to use a triangulation of quantitative and qualitative methods. Our methods have included presence, IMI and flow questionnaires, oral interviews and both video and system log analysis. As stated by Wagner et al. (2008) most researchers agree with Slater & Steed (2000) that presence has a subjective, psychological, as well as an objective, physical component. Consequently, evaluation methods range from assessing subjective phenomena (e.g., through questionnaires) to observing objective phenomena (e.g., by measuring bio-signals, performance times and so on).

Regarding to meta-analysis conducted by Dunser et al. (2008), only 10% of Augmented Reality research papers published by ACM and IEEE included any kind of user evaluation. In these evaluations objective testing was the most used evaluation method and formal qualitative analysis not as common (only 9 publications). They also found that evaluating collaboration between users has been quite underrepresented (only 10 publications). Augmented Reality research can be seen as a part of the wider Mixed Reality research field and the findings of Dunser et al.'s study can be probably applied to it as well.

Dunser et al.'s (2008) findings have encouraged us to develop further our multi-method approach, that was selected to encapsulate the multi-faceted phenomenon of Mixed Reality interaction, which includes not just the person using certain technology, but also her surroundings (the urban environment in our case) and the other people she is with while using the technology (who can be bystanders, co-users or users who don't use the technology themselves).

As a background we worked with theoretical work from EU projects iPerg and IPCity around orientation and interaction between users, devices, the environment, spectators and researchers. We also looked at how players focus on, act though or use artifacts as mediators (Norris 2008).

Our analysis method can also be labeled as *multimodal analysis*, which is an approach to representation, communication and interaction, and which looks beyond language to investigate the multitude of ways we communicate: through images, sound and music to gestures, body posture and the use of space (Jewitt, 2009).

From the different multimodal dimensions highlighted by Wagner (2009), we have focused in our video analysis process especially in the dimensions related to embodied interaction, as suggested by Dourish (2001):

- Gestures
- Gaze
- Body posture and movement
- Object manipulations
- Use of space

Our choice of focus in embodied interaction was selected for three reasons: Firstly, with Mixed Reality technologies, user is shifting constantly between the virtual and physical worlds. To properly analyse this multiple reality management and to understand all the factors that affect to it, we must go further from just interpreting the user experience afterwards through interviews or tests, to observing how users manage the technology and construct their experience from the both worlds with other people while it happens. Secondly, Mixed Reality technologies have not been investigated from this perspective before. Thirdly, from the very beginning of our field evaluations, using our grounded theory (see Glaser & Strauss, 1967) based approach, we have noticed that bodily interaction in the urban space plays a major role in the Mixed Reality experience: if one uses technology alone or with others or if the technology consists of physical objects that the users have to manage, just these factors can dictate how the Mixed Reality technology is used and experienced.

In the following chapters we will go through the different analysis techniques we have used in our research, and explain how they can contribute in exploring the Mixed Reality experience. We will describe the latest field trial with the *MapLens* technology, and will give an overview how the different research methods have been used together.

1.1 Evaluated technology

The technologies evaluated in this report have been developed in collaboration with WP4, WP5 and WP7. In this report we focus specifically on MapLens prototype and describe how it was evaluated in our field trials. In the interests of brevity, CItyWall evaluation will be discussed in more depth in public deliverables D4.4 and D7.4.

MapLens, which allows augmenting real paper maps with virtual information when viewed through mobile phone's camera, is based on the technology developed in WP4 in collaboration by HIIT/TKK, UOULU, UCAM and TUG.

The evaluated technologies will not be described in this document. For further details how the technology has been developed and how it works, see the public deliverables D4.4, D5.4 and D7.4.

1.2 Field trials, events and users

During 2009 we have organised multiple events and field trials around the showcase prototypes. Our evaluation was organised so, that different members of the IPCity project could participate in organising the trials, guaranteeing us as wide group of professionals from different fields as possible. Visiting researchers from FIT, TUG, UOulu, Nokia Research, New York University, University of Otago and HitLabNZ participated in planning and organising our field trials during the summer.

The showcase organised several workshops on evaluation with international participants. The showcase also succeeded in carrying out two field trials for the *MapLens* prototype.In addition the showcase had and is having a permanent installation for the Multi-Touch Display in Lasipalatsi, Helsinki.

Events and trials organised during the last year of the project are listed in Table 1.

Table 1. Events and trials organised by WP7 during the last year of the IPCity project.

Prototype	Date	Event/trial	More information available at	Participants
CityWall	Jan 1-Dec 31	City installation in cooperation with Cultural Office	http://citywall.org	Average 500 per week
MapLens	April 21-23	MapLens presented at FET2009 exhibition	http://ec.europa.eu/ information_society/ events/fet/2009/	800
CityWall MapLens	May-June	3 seminear series on evaluation		50
MapLens	August 6	Workshop on MR/AR examples and evaluation styles		15

MapLens	September 9	Workshop on interactional techniques in mobile virtual and augmented reality applications		15
MapLens	August 16	1st Field Trial		23
MapLens	August 23	2nd Filed Trial		14
CityWall	November 27	Workshop on Multitouch: Design Issues and Knowledges: Limitiations and Affordances	http://www.hiit.fi/~morrison/workshop27November.html	15
MapLens	September 22-25	Workshop on Environmental Awareness	http://ipcity.imagination.at/summerschool/	9

1.3 Dissemination and Demonstration

WP7 presented a paper on *MapLens* field studies in CHI2009:

> Morrison, A., Oulasvirta, A., Peltonen, P., Lemmela, S., Jacucci, G., Regenbrecht, H. and Juustila, A. (2009). Like bees around the hive: a comparative study of a mobile augmented reality map. In Proceedings of the 27th international Conference on Human Factors in Computing Systems (CHI '09) pp. 1889–1898.

This year WP7 has submitted two conference papers to the CHI2010 conference:

> Morrison, A., Lemmela, S., Oulasvirta, Schmalstieg, D., Peltonen, P., Mulloni, A., Regenbrecht, H., Jacucci, G. and Juustila, A. From Single to Multi-Lens Collaborative Augmented Reality on Mobile Phones. Submitted to CHI2010.

> Jacucci, G., Morrison, A., Richardson, G., Kleimola, J., Laitinen, T. and Peltonen, P. Worlds of Information: Supporting multiplicity at a public multitouch display. Submitted to CHI2010.

A journal article was submitted to the Personal and Ubiquitous Computing journal:

> Morrison, A., Lemmela, S., Peltonen, P. and Jacucci, G. Methods to Evaluate Pervasive Technologies: Games and Patterns of Play. Submitted to PUC.

This article was not accepted and will be resubmitted with changes in the HCI journal.

Also WP7 has written a book chapter about CityWall and MapLens in the Springer Series on CSCW:

> Jacucci, G., Peltonen, P., Morrison, A., Salovaara, A.,Kurvinen, E., & Oulasvirta, A. (in press). Ubiquitous media for collocated interaction. In Willis, K. (Ed.), Shared Encounters. Springer Series on CSCW.

WP7 organised several workshops during 2009. These are listed in Table 1.

2 Overview of General Analysis Procedure

2.1 Questionnaire Analysis

Immersion in virtual realities has bee traditionally researched through use of presence questionnaires. Lombard and Ditton (1997) have defined the feeling of presence as the perceptual illusion of non-mediation, which has three dimensions: spatial (the feeling of "being there" in a mediated environment), social ("being together with another") and co-presence ("being socially present with another person"). Traditionally, to study the different aspects of presence, presence questionnaires have been used for this. We selected to use the MEC-SPQ (Vorderer et al., 2004) presence questionnaire to investigate the spatial presence experienced by the participants in our trials.

As reported earlier by Morrison et al. (2008), we have looked to the work of Csikszentmihalyi (1990) on flow and optimal engagement to extend our evaluation methods. Flow is described as an auto telic state, where people lose track of time and any self-consciousness surrounding their activity, as they become so involved in an activity that nothing else matters. When people complete the kind of activity which has put them into the flow state, they feel much better about themselves and life generally. Activities may range from e.g. mountain climbing to painting. There are a multitude of activities that the work of Csikszentmihalyi (1990) has shown can produce this state in individuals.

The original concept of Csikszentmihalyi (1990) has been adapted to understanding flow in gaming by Sweetser et al. (2005), who have developed a questionnaire to measure the flow experienced in game like situations. In our work, we are using a similar approach to improve user experience on large touch displays and to investigate the user experience afterwards in our trials

A third questionnaire that we have been using to gather feedback from our field trials is the Intrinsic Motivation Inventory (IMI), which is a multidimensional measurement device intended to assess participants' subjective experience related to a target activity: interest/enjoyment, perceived competence, effort, value/usefulness, felt pressure and tension, and perceived choice while performing a given activity. The interest/enjoyment subscale is considered the self-report measure of intrinsic motivation. The device has been used in several experiments related to intrinsic motivation and self-regulation (e.g., Ryan, 1982, Deci et Ryan 2000) and was originally designed by Deci and Ryan, 1994.

The questionnaires are available at:
http://www.psych.rochester.edu/SDT/measures/IMI_scales.php

The three different questionnaire models have been used in combination to get a better understanding of the audience experience and how they engage.

The questionnaire data is analysed statistically, for example using software such as SPSS. When combined with demographic information, valuable observations can be made such as "females reported experiencing more spatial presence than males". These kinds of subjective measurements offer interesting insight how users experienced the tasks and technology they were trying out.

Questionnaires can provide the researchers with interesting information, but one has to remember that these are subjective and retrospective measurements: what users report afterwards might not be the whole truth what happened in the field from the beginning to the end of the trial: the actual questionnaires might be interpreted differently by each participant: for example, the word "often" might have a different meaning for user A than for user B. Also, especially the presence questionnaires can contain very abstract terms and concepts that users might understand differently, depending on their age and educational level for example. Slater et al. (2007) have argue that in general presence questionnaire data is treated far too seriously, and that a different paradigm is needed for presence research–one where multivariate physiological and behavioral data is used alongside subjective and questionnaire data, with the latter not having any specially privileged role.

2.2 System Log Analysis

In the first field research manual for IPCity, Oulasvirta et al. (2006) highlight the importance of gathering system logs to get an understanding how the system was used, which will provide context for making interpretations of the presence measurements and other observations. For a communication based application Oulasvirta et al. (2006) give an example what should be logged (presented in Table 2).

Table 2. What to log from a communication based application.

• Average # of packets sent per day • What time of day was the system accessed • How active were different users in accessing the system • What is the average amount of packets per day • Distribution of media taken by the different phones • How many replies and how are they distributed among different users • When were replies made, time of day • How long time after a piece of media was uploaded was it commented • Average # of packets sent per day • What time of day was the system accessed • How active were different users in accessing the system • What is the average amount of packets per day	• Distribution of media taken by the different phones • How many replies and how are they distributed among different users • When were replies made, time of day • How long time after a piece of media was uploaded was it commented • Who replied/viewed a message, the author vs others • What is the average length of discussions (number of comments, length of an individual comment) • Distribution of different media (audio, video, photo) • Turn-taking length: how long between replies to a comment • Calling others before or after using the system. This is important for understanding the role of the system for coordination. • How many times were messages created/viewed when others were present vs. not?

The logging needs depend heavily on what kind of system one is evaluating. When evaluating a multi-touch screen for example, the first three items on the list are important, but others are probably not: in this case one would need to focus on how many users are at a time at the display, how many hands they use, what kind of gestures are being used and so on.

Good logging can reveal not only interesting patterns of use and give perspective to the questionnaire analysis, but also act as a filter for the qualitative analysis process: with good logging we can identify the most interesting sessions of use for more in-depth qualitative analysis. This is very useful, if we have hundreds of hours of video data gathered and only limited time to analyse all this material.

2.3 Video Analysis

As noted by Pink (2007; Ref. Wagner 2009) video recordings account for the situatedness of the visual, temporally and spatially, with respect to the environment; they make it possible to examine the gestural and scenic details of how people interact. Video analysis is a good tool for catching the different aspects of embodied interaction that other methods cannot capture, but it is also challenging to do reliably and quite time consuming.

2.3.1 Data collection

Collecting data by video recording in field trials is harder than it sounds. It is very easy to get distracted and target to wrong things, it is impossible to get everything on tape. If video recordings are the only form of data gathered and one is shooting subjects on the move, doing bottom up grounded theory (Glaser & Strauss, 1967) based analysis comes quite hard, as one has to decide beforehand in what to focus: the more clearer and focused the research questions are in the beginning, the easier it is to do the actual recording (and the analysis in the later stages). With static installations where the subjects don't move as much and you can cover the whole "interaction area" with cameras, you can do more explorative type of research to just see what happens when people encounter the new technology.

Another difficulty in video data collection is when to record: if you record an installation 24/7 you will end up with hundreds hours of data. With careful planning and doing demos of the actual trials one can save hours and hours of time in the analysis phase. Also proper logging of the system can prove extremely useful in this sense as discussed in chapter 2.2. And when on the move, one has to consider the battery life of the cameras. What could be more disappointing for a researcher than to find out that you have recorded hours of meaningless chitchat which have eaten all your camera batteries when something actually interesting starts to happen? And while the researcher is changing the battery she might also miss something essential that the users do.

The selecting and cutting of video material is done in what Laurier et al. (2008; Ref. Wagner 2009) describe as "forming the film as an object out of the materials that are there" in many cycles of previewing and reviewing, making visible what we think are relevant instantiations of participants' co-constructing the experience. This cutting process is already part of the analysis process, which general guidelines will be described next.

2.3.2 General video analysis process

In this chapter we describe a general video analysis process, which is suitable for most purposes when evaluating use of Mixed Reality technologies.

We have used successfully an analysis process based on "constant comparison analysis" (Glaser & Strauss, 1967), where codes emerge inductively, or through the data directly. In this process, at least 2 researchers go through the analysis process together. This way interrater reliability can be achieved, which is the extent to which two or more individuals (coders) agree while doing the analysis. Interrater reliability addresses the consistency of the rating system, which can be ensured if multiple people code the same data and agree with each other's codings.

The steps in the video analysis process are described in Table 3.

Table 3. Video analysis process steps.

1. Initial research questions are formulated by the researchers
2. The researchers watch the entire video first separately, and make field notes of interesting points segments (relevant clips) that they would like to explore further.
3. The researchers meet with their notes and negotiate which of the selected clips are the most relevant for the research questions. Then these relevant clips are sorted and uploaded to a shared folder from where all persons involved in the analysis can access them.
4. The researchers watch each clip separately and come up with a list of codes associated with each clip.
5. Achieving interrater reliability: The researchers come together again and compare the codes assigned to random selection of the video clips. If differences are found between the researchers' coding, these are discussed and the coding scheme is modified if necessary. The researchers repeat this step until no differences between

the coding results are found. When resolving differences between coding schemes, think about the following: What are the relationships between the codes? Do the codes really capture what's happening here?

6. After researchers have agreed with a common coding scheme, the researchers will then recode the videos together based on this new scheme.
7. Finally, they work together to integrate the themes that emerge and finalize or refine theory based on the final coding and analysis.

2.3.3 Tools for video analysis

There can be found numerous commercial and free transcription systems to help in the video analysis process. Two simple freeware programs can be found from these websites:

http://videonotetaker.sourceforge.net

http://www.dvcreators.net/qt-movie-notetaker/

A suggestion for a commercial solution could be ATLAS.ti for example, which allows an easy way to manage the codes you have created and build diagrams etc. based on them. But most of the video analysis tasks can be easily done also with regular video playing software such as VLC or mplayer and using a spreadsheet program for the codes on the side.

2.4 Interviews

All the earlier methods of gathering and analysing data rely heavily on the researcher's interpretations of why the users used the system under evaluation the way they did. This kind of analysis has always the risk that user's own intentions and meanings for their actions do not get exposed, and something valuable might be missed in the analysis process. Therefore it is a good to hear the users to explain their experiences in their own words. This way the Mixed Reality experience can be framed yet from another angle.

2.4.1 Interview techniques: semi-structured, cue based and researcher interviews

In *semi-structured interviews* the users are asked a set of predefined questions in a flexible way, allowing new questions to arise in the interview as a result of what the interviewee says. This way all the users are given the option to describe freely their experience related to the topics that interests the researchers.

Sometimes the interview questions can be too abstract or complicated for the interviewee, or she can feel pressure to give what she feels is the "right" answer for the question. A good tool to overcome these problems, and to make the interviewee have more freedom to explain her experience, is to have also cue-based parts in the interview. This kind of *cue-based narrative interview*'s aim is to get the informant to recall *real, actual episodes that happened and to tell them in her own words*. In the interview the interviewee is presented with cues such as video footage from the trial that helps her to recall the experience. Oulasvirta et al. (2006) describe the general approach of this technique as follows:

> Using a (color) print with one screenshot of the technology in question on each page, point a feature/data item/object and ask the interviewee to explain what aspects s/he used, what it means and then to tell 1-3 episodes for that. Because the use of these cues might have been quite uncommon, it is even more important not to give up as an interviewer but to pressure even more to find out even the rarest and most marginal use cases.

Another technique is to give the actual technology back to the user and ask her to show how she used it for different tasks. Researchers can also this way capture some of the embodied interaction and compare that to the ones they have observed on the field.

When dealing with data gathered from multiple sources, it might be sometimes also useful to interview not just the users but the also the researchers that were observing them on the field. This way it is possible to get a wider perspective of the user's actions in the field than can be witnessed from video recordings, which can focus only on a single thing at a time.

To increase recall, during these interviews one can also make the interviewee watch video footage captured during the trial. For more information about this video (cue) based recall technique see Costello et al. (2008).

2.4.2 Interview data analysis

In the actual analysis part, the interviews are first transcribed into textual format. The amount of detail in the transcription can vary based on what one is looking from the data.

After the interviews have been transcribed, the data is content analysed in a similar way that as with the videos (see chapter 2.3.2 in this document). The idea is to come up with a list of codes that describe the experience. The abstraction level of the code is also dependent on what one is trying to dig up from the data. This excerpt from our MapLens 2009 study (Morrison et al., 2009), which presents an coding system that focused on how users described their experience with the mobile AR system (M) in comparison to digital only system (D):

> In the transcriptions of our interviews, we searched for recurrent adjectives in the participants' descriptions of their experiences. We found M users made 11 mentions of the word *stability* (and 0 with D). For example, "You need to be quite accurate; you need to be *stable* and you need to get the camera into the right position." Six M users described the trial as *easy* compared to 25 instances of *easy* being used by D players. Here too, we find M teams more challenged by the technology: "At first it was difficult to find these dots. Maybe it was because we were not able to keep our hands stable enough. But after that we catch the red dots by using the square."

In some cases it might be useful to lift the abstraction level of the coding, for example one might create codes labelled "it might be just enough to code "positive experiences" and "negative experiences". Of course the same data can be analysed at multiple levels at the same time.

To achieve interrater reliability, it is wise to follow the same guidelines as in video analysis: multiple researchers code the same episodes and then compare their results, after which the coding scheme is modified if necessary. These steps are repeated until researchers agree that they are coding in the same way and reliability has been achieved.

3 Evaluation Case: MapLens

Following previous studies on collaborative use in mobile Augmented Reality, we set up a field experiment to better understand differences in collaboration and tangible Mobile AR device use in urban environment in August 2009. In this field trial participants used MapLens (see Figure 1), an application on a mobile phone that works like a magic lens over a paper map, which provides an additional layer of digital information to the view seen hrough the mobile phone's camera.

Figure 1: The MapLens application showing a live video of the paper map underneath, augmented with icons and labels registered to map locations.

Our study was the first study of its kind to synchronously trial multiple, single and shared users and mobile devices in the field. The three configurations were: solo users with one device; a team of three sharing one device; a team of three with each one device. Each configuration completed the same game tasks in the same given time. We found that solo users could complete the game tasks in the given time therefore shared use as not required. However, in teams with more devices, the devices were used in a more expansive way. We observed divergent roles emerging and that the teams still decided to share only one paper map. We also noted that teams largely stayed together to complete tasks, despite it was not essential to complete the game. In teams sharing the device, looking at and pointing at each other's screens and the map beneath, occurred more than in the teams where everyone had their own device. The findings of this study have been reported in more depth in a paper submitted to CHI2010 (Morrison et al., 2009b).

In this section we will look into more depth how the evaluation of the field trial was done using multiple and both quantitative and qualitative methods following the procedures and methods described in earlier sections. This section's purpose is to show how these methods can be triangulated in practice and what kind things has to be considered especially when evaluating Mobile AR applications.

3.1 The trial

3.1.1 Research Questions

Before the trial we held long discussions and brainstorming events with TKK researchers and visiting researchers from HitLabNZ, UOtago and TUG, how we could extend our evaluation of the *MapLens* system from last year's trial and how the new AR features would be best trialled. We discussed many things, some of them were as follow ups from mapLens1 trial August 2009 to prove the findings we found there without limitation. These possibilities included questions listed in Table 4.

Table 4. Research question options.

- Changing the role of the physical objects—take away/ add objects
- Using a non-AR game as a comparative—just using a paper map with the same game

- Comparing different map sizes between groups (see also Rohs, 2009)
- Comparing size of the groups
- Adding one extra play component to test for anarchic play state and radical behaviour
- Analysis could look at Suchman's notion of situated action for the interaction work between team players

Other considerations included:

- One digimap session with equal gender mix
- Ensuring devices swapped between users equally as part of the game tasks
- Comparative testing of tracking with new and old system—lab or in field, located but without tasks
- Using more experienced researchers to ensure accuracy/ even-ness of the reporting

As addendum and background related material that contributed to the research questions we considered the following issues

- Will improving performance-type issues (delay, fuzziness, difficulties reading maps, difficulties with the interface, difficulties reading the icon information, providing information on 'you are here') impact on how MapLens users collaborate, common-ground and negotiate when using the device? Tested with condition Group 1, 3 and 4a
- What available features do our users use the most? And what do they do with them and why? How long for? (Improved logging shows this)
- How task-orientated/ distracted/ playful are our players? How rule-bound is their play? How anarchic is their attitude to tasks and the games? (Supports research we do around game tasks and kinds of behaviours they elicit)
- How liberating can games be for users? Can users become immersed in game world (magic circle) and forget usual inhibitions?
- Situated action: unpack what it is we call interaction activity/ interaction work. Devise a common-ground language to discuss this interaction work and embodied activity
- Prove collaboration and embodied interaction happen (or not) regardless of number of devices or regardless of AR

As reported in Morrison et al. (2009b) in this experiment we decided to test three conditions:

1. Three devices and three maps in a team of three people, denoted as 3/3
2. One device and one map shared in a team of three people, denoted as 1/3
3. One device and one map for one person solo, denoted as 1/1

See Figure 2 for a graphical representation of what conditions were trialed which day. We decided to print the map larger on paper (not foam core) like a usual paper map that folds up, so we can run single use, multiple use and shared use according to the numbers at each trial.

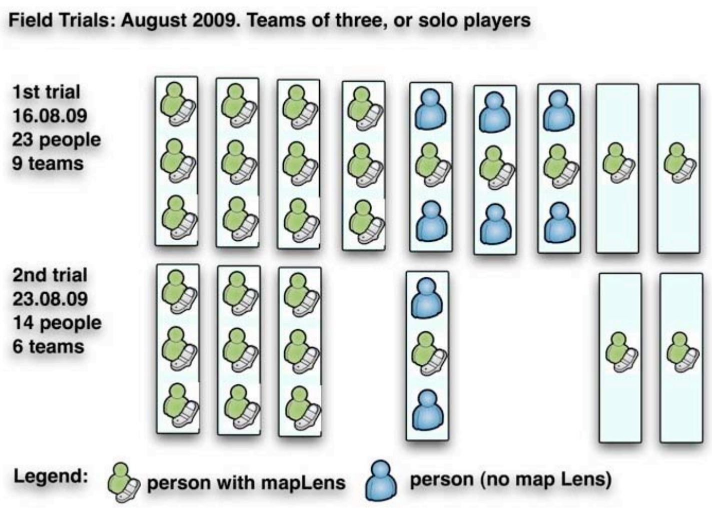


Figure 2. Configurations of players on the 2 trials days.

3.1.2 Design of the trial

The trials were designed as location-based treasure hunt games in the Museum of Natural History and the green areas of the city. Unlike in earlier work in which environmental games have been largely narrative-based (Klopfer, 2008) the goal of our game was to connect players with urban nature by giving them a new kind of experience of the city. The goal was to make their connection to urban nature and place to endure beyond the more artificial environment of the game. As such, our aim was to re-position physicality at the core of our players' AR experience by including many artifacts, and designing the game and tasks to remind the participants of their own selves, interacting within the physical world (Merleau-Ponty, 1968).

3.1.3 Briefing the researchers

Each team had a researcher video-record and observe them in the field. As there was 15 teams in total, we needed to brief all the researchers before the trials so they would know what to look for. The researchers were briefed with the following topics:

- How to work with the video camera (and prepare ahead)
- How to work with *MapLens* (what to do if it crashes, battery runs out etc.)
- What are our research questions
- What to focus on when videoing the participants

The instructions for our research team is listed in Table 5.

Table 5. Instructions for researchers.

* Synchronize ALL clocks: personal, those in the mobile device/server that logs interactions, videocameras, and audiotapes. **What are we looking for?** 1. Player-Player Interaction 2. Player-devices Interaction 3. Player-Environment Interaction 4. Player-spectator Interaction 5. Player-Game Management Interaction
What else do we focus on?

1. Gestures we observe: Iconic gestures represent something, such as motion, the size or shape of an object, Deictic gestures point to an object or place or in a direction. To space: immediate- close/ far Between the players (interactional space) **2. Gaze** we observe: Follow a gesture Direct attention Body follows gaze (move towards what looking at)
3. Body Posture and Movement we observe: Orientation within space: e.g. around card map, a device, towards real environment, other artifacts Orientation towards each other Orientation towards the environment
4. Object Handling we observe: Attending to and acting to the thing-focus Acting through the thing (extension of ourselves, unaware) Thing as Mediator—use it, aware but not focus (like common ground) (Wagner, 2009, *IPCity Guidelines for the set up and analysis of trials*. pp. 7-9)
PRINCIPLES VIDEO Capture: 1. THE CAMERA AS A SPOTLIGHT. Think your video camera as a spotlight. Although you're there, witnessing with your all your senses and wide field of vision what's happening, your video camera picks up only a tiny part of that. Camera is a sampling device, you're the sampler. 2. CAPTURING HCI. We are studying human-computer *interaction*, which means that we have to capture the user and the device. However, this is a special case: there may be more than one user AND the device has a referential relationship with the surrounding environment (because it's a map!). The moral is: 3. BE PERSISTENT. Aim at 100% quality. You have to stay sharp all the time. Do not give in and think that there's enough data already.
GUIDELINES A) PRIORITIZE THE USER(S). Often when the subjects are talking about or pointing at or orienting to some STATIC object in the built environment, as they will be doing many times, it may not be THAT important to keep the video on that object for a LONG time. Rather, prioritize the users. In many cases it's obvious from what we know about the task and the spot the users are what they are talking about. Try to capture what the users are doing together and keep in mind what they pointed at. If the situation is brief, you can capture that object after the users have stopped talking. B) *NO* TALKING WITH THE USERS. Your task is to record the interactions AS if you were not there. You are not supposed to talk with the users or answer their questions. However, if the software crashes, your duty is to fix the problem. Intervene, don't leave them struggling with it. Remember they have to learn a lot in their briefing session so they may forget some things. ALSO if you see that the participants are under-using or incorrectly using the

application, PLEASE assist them. For example: “do you remember you can use 1 to stop seeing all the uploaded photos and press 1 again to turn them on?” This is the only interaction that is allowed with the users. The participants usually react to your presence (being videoed) by making remarks/jokes about the camera or "acting" for the camera. You should not care about that but stay neutral, eventually this will all relax. Film them in the museum so they get used to this (and used to searching).

C) YOU FOLLOW THE USERS, NOT THE OTHER WAY AROUND. You should not imply or hint where to walk. You should not hint what the correct answer is. If you don't believe, read the classic story of the Clever Hans:

http://en.wikipedia.org/wiki/Clever_Hans :)

D) EFFICIENT SHOOTING DISTANCE IS 1-2 METERS. The only way to make sure that you can fully capture interactions is to stand close.

E) DON'T RECORD THE BACKS OF THE USERS. It's important to see where they're looking at and what they're pointing at. The optimal angle is a little bit to the side of the user.

F) DON'T BLOCK THE VIEW OF THE USERS. Don't stand in front of the users. Remember: You're not there.

G) MAKE THE USERS COMFORTABLE. This is not ethnography. We don't have the luxury of spending years with the participants and make them comfortable with your presence and the camera. Therefore it is necessary that you spend some time in the *beginning* introducing yourself, maybe even cracking a joke / ice breaker. But only in the beginning when you meet, not when the action is on.

H) CHECK YOUR CAMERA. Your responsibility is to collect data. Sometimes technology disagrees with. If you notice that too late, then we lose a whole session. So, just before embarking, check your camera. And do it again when it's safe.

I) YOU MAY HAVE TO RUN. If the participants are walking and doing something with the materials during walking, you may have to run, especially at corners. (You may think yourself as a satellite on the sky: your path is always longer than that of the planet's).

J) BE CAREFUL WITH THE SUN. You must know this, but just to remind you.

L) FAMILIARIZE YOURSELF WITH THE CAMERA. How does it operate? How do you change the tape/battery? Can you index events (useful!)?

M) KEEP AN EXTRA BATTERY + TAPE WITH YOU.

N) ENTERING A SCENE. When the users enter a new scene, you may take a "panoramic view", but ONLY if they are not interacting with each other. Remember that you can also do this later on. Otherwise, keep the focus on the users.

This time we decided ahead of four places we would ensure that the researchers videoed with close attention the participants—four places where we knew that would need to orientate to the environment, to each other and through the device find their current location and where to go next. This was to ensure we got the footage we needed.

3.1.4 Briefing the participants

We had a total of 37 users, 19 females and 18 males between the ages of 14 and 44. The user group consisted mostly of expert users. Before the game the users went through a briefing session, where each team was handed their MapLens devices and a kit bag that contained a clue booklet and material needed during the game. The participants also were introduced to the researcher that was going to observe their team and who then showed the team members how to work with the technology.

The purpose of the briefing session was to make the participants

- understand how the game worked
- understand how the technology worked
- understand the role of the researcher and the video observations

The participants viewed a presentation, which explained 1) configurations of teams and which teams were with which researcher (see Figure 3 left), 2) the purpose of the game and how to approach it 3) the clue booklet (Figure 3, Figure 2 right), 4) how to boot the application and work GPS, 5) how to use the interface (see Figure 4) and how different MapLens features worked (see Figure 5), as well as 6) the general running order of events, handing out of prizes etc.

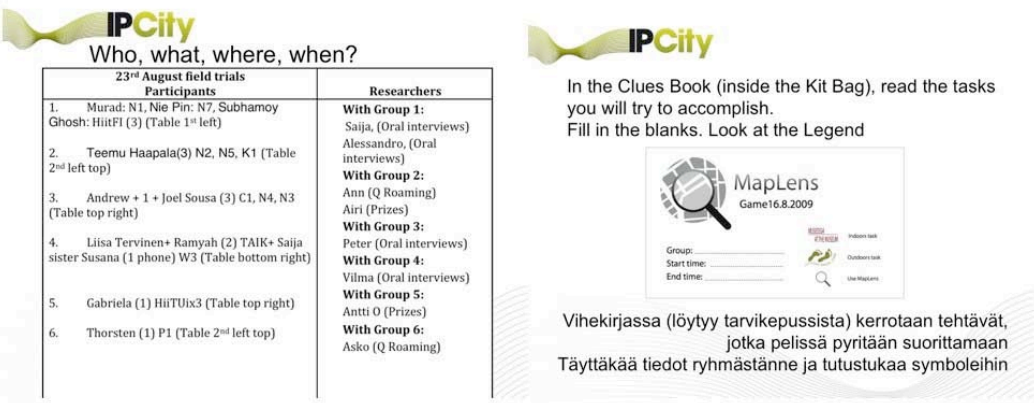


Figure 3. Left: MapLens introduction slide 1. Right: MapLens introduction slide explaining the clue booklet.

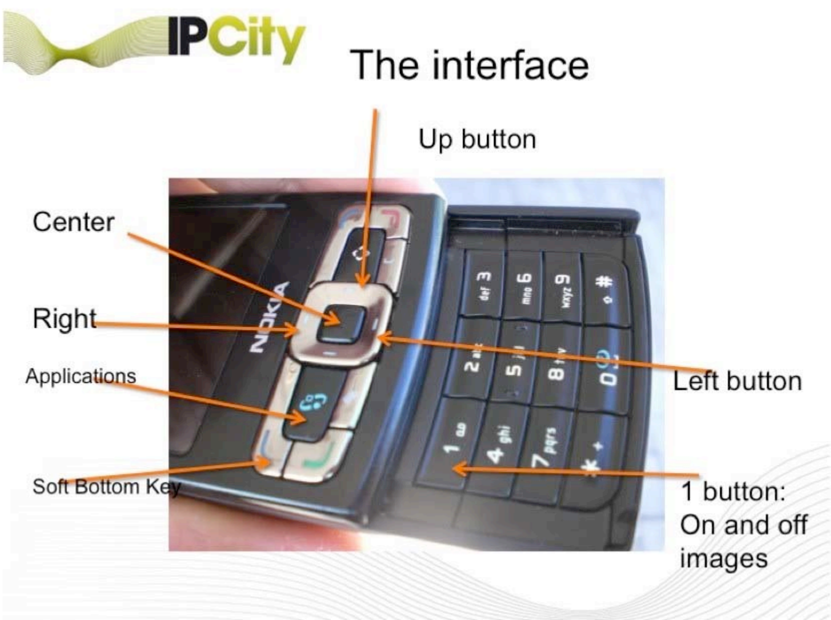


Figure 4. MapLens introduction slide explaining how to use the device.

Figure 5. MapLens introduction slides explaining how different features work.

At the end of the intrudction session the participants they received final instructions how to start playing the game (see Figure 6).

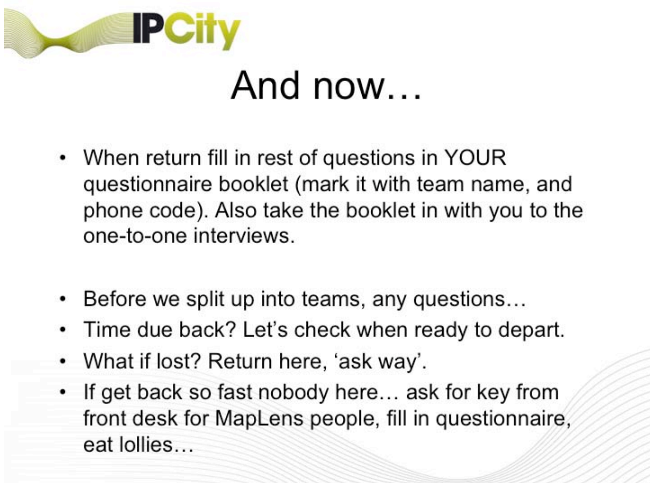


Figure 6. The final MapLens introduction slide.

3.1.5 The game

The game began after a training and introduction session at the museum. The teams completed six tasks at the museum without any assisting technology, but when planning their outdoor activities and routes, MapLens supported them. MapLens showed images as clues, like an image of a recycling point or a statue, which guided players to locations, where the tasks could be conducted. Not all outdoor tasks necessarily required MapLens use for successful completion, as we wanted to direct attention to physical aspects of the environment. Many tasks included photographing, and all the photographs taken during the game were shared through MapLens between all the players, providing the players a feeling of stronger social presence with the other groups.

The experiment lasted for an hour and a half. The weather was roughly the same for both days, cloudy and windy.

3.1.6 Data collection

Before the game, participants filled out forms including questions on demographics and experience with technology, use of maps, knowledge of environmental issues and Helsinki centre. Throughout the game, one or two researchers taking video accompanied each team, having been instructed to focus on particular instances of use and types of interactions.

After the game, participants completed shortened versions of a MEC Spatial Presence Questionnaire (MEC-SPQ), GameFlow questionnaire and an Intrinsic Motivation Inventory (IMI) to gauge reactions to the game. Participants were also interviewed with in a semi-structured interview, where they were also asked to show how they had used *MapLens*, which was then videoed.

After the trials, videos were cut into manageable chunks focusing on activity around the MapLens system and tasks. The footage from two teams, a shared device team (1/3) and a solo team (1/1) was excluded from the analysis due to technical failure. Then each researcher that had observed the teams in the field participated in a 30-40 minute semi-structured interview with the core team of researchers. Next, we will describe the analysis process of the data in more detail.

3.2 The Analysis Process

After the trial, the data we had collected included:

1. Demographic questionnaire data
2. MEC-SPQ questionnaire data
3. GameFlow questionnaire data
4. Intrinsic Motivation Inventory questionnaire data
5. Game team video observation data
6. Participant interview data
7. Participant *MapLens* video demonstration data
8. Researcher interview data

Post-processing tasks for all this data included entering the questionnaire data into spreadsheets and cutting the videos into episodes that could be analysed. The analysis process in whole included the following steps:

1. Post-processing the questionnaire and video data
2. Game team video analysis pass 1: First pass with the videos, formulation of initial categories
3. Researcher interviews with videos as cues

4. Game team video analysis pass 2: formulation of 52 item code list
5. Game team video analysis pass 3: coding the instances of the activities in the 52 item code list into spreadsheet
6. Game team video analysis pass 4: drilling down with the video analysis, focusing on things revealed by the coding (team roles for e.g.)
7. Questionnaire analysis with SPSS
8. Participant interview analysis (coding into spreadsheet, creating categories, counting instances)
9. Participant MapLens video demonstration analysis (coding into spreadsheet, creating categories, counting instances)
10. System Log analysis
11. Triangulation of data: mapping the results from the different analysis steps together to find and make sense of patterns that cross match results.

In the following subchapters we will go through the analysis steps by method used.

3.2.1 Questionnaires

Before the game, participants filled out forms including questions on demographics and experience with technology, use of maps, knowledge of environmental issues and Helsinki centre. After the game, participants completed shortened versions of a MEC Spatial Presence Questionnaire (MEC-SPQ), a GameFlow questionnaire and an Intrinsic Motivation Inventory (IMI) to gauge reactions to the game (Vorderer et al., 2004), (Sweetser & Wyeth, 2005), (Deci & Ryan, 2000). For Presence questionnaires we measured concentration, errors, activated thinking, and imagining space. For IMI questionnaires we measured interest/enjoyment, perceived competence, pressure/tension, and effort/importance**.** For Flow questionnaires we measured challenge-skills balance, clear goals, concentration on task at hand, and sense of control. For social presence we added questions under development and validation through the EU funded IPCity consortium that investigates presence and interaction in urban environments. These provided us with quantitative data about the user experience, as did the in one-to-one semi-structured interviews that followed and are discussed in the next sub-chapter.

We looked at the questionnaire analysis results as an additional resource that was used as support for the video analysis. In our opinion the default presence questionnaires are sometimes too abstract and sometimes too specifically designed for virtual reality research to be used as such in Mixed Reality research, where the experience is created through technologies that vary greatly how they are used. We translated them to a ‘common-sense’ language, still retaining the original meaning and with consistent meaning for their translation into Finnish (and making sense in that culture). We did this with four researchers: one presence questionnaire expert, one evaluation expert, one mobile technology expert and one ‘using mixed methods for evaluation’ researcher. This process forced lively debate and took the most part of one day, with further follow-on conversations over email and in the translation process with other evaluation experts.

The English version of the questionnaire can be found from Appendix 1.

3.2.2 Interviews

User Interviews

In the semi-structured oral interviews after the finishing the game, the participants described their experience, highlighting aspects that had caught their attention in the game. These are the example questions asked from the participants:

Q1. How did you use the MapLens, can you show with a phone and a map?
Q2. Did you know beforehand your team members? What relationship—friend, colleague, boss etc?
Q3. Did pointing help you complete the map+phone tasks?
Q4. Did talking with the others help you complete the map+phone tasks?
Q5. How was the experience?
Q6. Which parts did you take more time with? Which did you enjoy most? Which things related to game or technology, were you thinking more about /played more with / returned to or engaged most with.(Choose which part of the question is appropriate to your interviewee)
Q7. And then if the user said something interesting I would ask more about it, but letting the user speak as freely as possibly.

All answers were recorded with a digital sound recorder, except Q1, which was also recorded with a video camera: participants were asked to show with MapLens how they had used the device and this was then videoed, allowing us to see them using the device in the more controlled environment of the museum (even lighting, no wind etc.) Also this session acted as a cue for the user to go back to the experiences she had encountered in the trial with the technology. The participants were asked to think aloud when doing this, so that the researcher could pinpoint important points that might have been otherwise missed in the interview.

Data post-processing tasks included transcribing these user interviews and translating them into English for further analysis. Also, data from the pre-phase forms and the post-trial questionnaires were entered into spreadsheets, and videos were cut into pieces where activity around MapLens tasks occurred.

The actual interview analysis included coding the interviews into a spreadsheet. In addition to the questions listed earlier and demographic information such as gender and education level (see the questionnaires in Appendix 1), we also coded things related to categories listed in Table 6.

Table 6. Categories coded.

• Map	• Roles
• Multiple thumbnails	• Pointing
• Enlarging pics	• The experience in general
• On the move	• Most time consuming
• You are here	• Most enjoyable
• System in general	• Most engaged with
• Talking with others	

The interview coding data can be found from **Error! Reference source not found.** (Sheet1: Data): double click the figure to see the whole table.

This coding then could be then analysed by counting the instances of occurrences in each category for each three condition. The occurrences could be then categorised further. For example for the category “the experience in general” we counted:

- 10 occurrences for the shared condition, of which 8 was positive, 2 neutral and 0 negative

- 3 occurrences for the solo condition, of which 2 was positive, 1 neutral and 0 negative
- 13 occurrences for the three devices condition, of which 13 was positive, 1 neutral and 2 negative

From this we could conclude that mostly users gave positive feedback about the experience in general while only the three devices group had something negative to say. Then we could drill down what negative things the three devices group encountered and start thinking why it was like this for this condition? Taken out for the purposes of anonymisation. The videos from the interviews (question Q1) were analysed separately by one researcher.

Researcher interviews

In the week following the trial each researcher, participated in a 30-40 minute semi-structured interview with the team of 3-4 core researchers, to obtain a richer overview of how teams interacted, e.g., how roles were formed, when discussions happened, how map and device shared, typical ways to gesture and point, and ways teams interacted with other teams, spectators and researchers. We also wanted to allow time for the researchers to reflect on what they had witnessed, as we had the immediate responses of the participants, as well as immediate footage at the trials.

To help recalling the important events, we watched together the video footage of the group the researcher was observing in the trial, adapting a video-based recall technique standardly used with participants (Costello & Edmonds, 2007). These videos acted as cues for the researchers to explain the interaction and events she had witnessed during the trial. This video based interview technique is discussed in the following chapter on video analysis.

Our process in research interviews was this:

1. We interview researcher with grabbed footage and get most information
2. Stay as a group with the video footage straight after the researcher interview and revisit for the last three stages of the process.
3. Collate what we find on the spot
4. At various stages in the process we need to total the similarities and differences in what we are finding, so remember this advice: “One way, start with characterization of typical and marginal use types *per group*. This is where we aim for in first pass.”
5. We can then plumb these more with in-depth of a sample, e.g. a group of 3 devices, a group of shared device users and a solo user gropu doing same task or where lots of activity for each group.

This way we focus our coding when we do the interviews. In Table 7 you can find a list of examples we asked from the participants.

Table 7. Examples asked from participants.

1. How roles formed? Who did what? phone-map-cluebook-bag. When did they switch roles? At what stage roles defined? System in place?
Example note: “Two lead roles, girl passive as late call in, guy in black took first leading and expert role, then two boys co-lead or battled for leadership. Both used maplens concurrently on map. Roles defined from museum outwards, used on ground when started, rolling map and dropping down, rolled map inwards system. Used ad hoc and on the move batteries, then it went awry because they separated and there on in stayed together. Grey guy used pen for clue book.”
2. Main person using *MapLens*? (in what circumstances did they switch?)
Example note: “Two leads, who got there first began it, and then e.g. taking photo”

3. Pointing gestures (on map and environment, screen, clue book; with pen and finger?)

Example response: “on screen many, on map many, not with pen, phone to point and phone to circle iconic gestures.”

4. When and what kinds of gesturing happened? Over map? At environment? When difficult?

Example response: “Some gesturing to the environment, and gestures with map rolled to the environment.”

5. Sharing device?

Example note: “All 3 person users using simultaneously on two occasions. Sharing, and looked through others, pointing on the others device, parallel use—are you getting this? Communicating while use in parallel.”

6. Other use of the device?

Example response: “Only for photographing, pointing iconic circular gestures No browsing, not great use of other photos (check)”

7. Map/s, Switching? (Whose map using)

Grey guys map used mainly. But black guys also. First all maps out., and then who has the map out first, although she never puts

Example response: “Girl when solo, black guy when couple, grey guy when all. Leadership and map related. People augmenting and map gets taken away by map owner (just like phone)”

8. Use while walking?

Example response: “Tried to use, and she tried a few times---map in held bag and tried to use it. Up and again and down again with girl trying to use in parked mode, not able to use while standing as not steady enough? (wind) map itself.”

9. Two-handed or one-handed use of the device (change over time?)

Example response: “Two-handed for clicking through images, enlarging etc. and one handed for roving the map, standardly in one hand horizontal use”

10. What was the alignment of the phone (e.g. vertical/horizontal, near/far from the body, can others see the screen etc.)

Example response: “Vertical or other while walking and horizontal on top of map”

11. Did they switch attention between the map and the mobile device? (Between phone+map and environment / From map to environment)

Example response: “When searching something from environment looking around, she going from mapLens to environment, often from all perspectives.”

12. Did they interact with spectators? With you as a researcher? With other teams?

Example response: “Interacted with researchers, reverse roles and took photos, interact slightly to Thorsten, did not interact with spectators and were self-sufficient”

13. Other comments?

Example response: “When split and did ad hoc activity at batteries, did not work out. Everybody carrying map and phone out all the time. Mediated image through feet in grass, looking through all cameras.”

Rationale

We reasoned that after going through this interview process we would have much information and should only need to revisit the videos for the following topics::

1. Object handling
2. Body Gaze (body follows gaze) (BG)
3. Gestures: Iconic and Diectic (or may be covered 4 + 5)

We were completely incorrect with this. We had a general impression of differences, but we did not really get enough information on how people used devices differently until we began actually counting specific instances of different types of activities. It was very hard to get beyond the personalities driving the teams and we knew we needed to do some analysis specifically with this. Who uses the phones and how (nature of the collaboration) depends on team composition and personalities. With MapLens1, we saw team personalities impacted use but had no evidence, so for MapLens2 we ensured we could prove impact, adding to an understanding of situated collaboration around mobile devices.

3.2.3 Video

Data collection, post-processing and preliminary analysis

Throughout the game one or two researchers taking video accompanied each team. Our researchers were briefed to record for the entire 90 minutes, but to focus on sharing, turn taking and object handling of the device, and instances where

1. The participants used *MapLen*s in the museum
2. The participants used *MapLens* outdoors
3. The participants were developing or changing strategy
4. The participants were working on a pre-selected task that required extensive *MapLens* use.

After the trials the video data was uploaded to a shared server, from where each researcher could access the files. The videos were then encoded into format that all researchers could access.

At this point the lead researcher went first time through all the video material and did a preliminary summary of observations of each team. An example of this kind of summary can be found from Appendix 2. Each summary contained only new kinds of activities found from the group observed.

Watching the videos with the researchers

After the preliminary analysis of the video material the core researcher team (four people) negotiated what they thought that was the relevant material in the video data. The summaries done by the lead researcher acted as the starting point for this work.

Each group's video footage was watched by the core researcher team together so, that the researcher who was with the group was also present and explaining what in her opinion happened in the video. This part acted at the same time as the video based research interview and the first pass of the video analysis.

At this point the researchers started forming the initial codes, mainly based on the multimodal dimensions discussed in chapter 1, but also on gaming related categories highlighted by the IPerg (2005) project. The coding system is discussed in the next subchapter.

Coding system

In our video analysis process we focused especially in the dimensions related to embodied interaction (Dourish 2001) and how players focus on, act though or use artifacts as mediators (Norris, 2004, iPerg, 2005, Wagner 2009). This focus originated from our research questions and the theoretical framework discussed in chapter 1. Based on that, we created initial categories (codes) of things we were looking from the video (see Table 8).

Table 8. Initial categories for coding.

Game Date: Team: Number of Devices:
1. Player-spectator Interaction (PS)
2. Player-researcher Interaction (PR)
3. Teams to other teams (TT)
4. Map use only one at a time? Record switching+map activity (MU)
5. Object handling: where the thing is LARGELY the mapLens in tandem with the map PLUS occasionally aspects of the kit.
• Attending to and acting to the thing-focus (TF)
• Acting through the thing—unaware of it in hand (extension of ourselves) (TA)
• Thing as Mediator—use it, aware but not the focus (TM)
6. Pointing:
Map (PM) Environment (PE) Screen (PS) Clue book (PC) Pen (PP)
7. Body Gaze (body follows gaze) (BG)
8. Gestures:
• Iconic Gestures (e.g. motion, size or shape of obect))IG)
• Deictic Gestures point to an object (DGN)Near or (DGF)Far or between players—interactional space (DGI)
Be aware we may need to include observations (MapLens1) For example:
- Pointing gestures (on map and environment) - Two-handed or one-handed use of the device
- Alignment of the phone (e.g. vertical/horizontal, near/far from the body, can others see the screen etc.) - Role switching and negotiations - Body postures and attention switches between the map and the mobile device

After the interview session and initial coding, the core researchers discussed their observations. These observations were mapped to the AR features of the system, listed in Table 9.

Table 9. AR Features of *MapLens*

A: Features Used	B: Improvements and new features
Green circle (you are here)	Taking images offline and browsing while walking
Selection viewfinder (red square)	Stability of use, so no issues with hand shake
Icon information (clues)	System more robust, so ease of use with ML system
Camera information (photos)	Interface gives more feedback

Thumbnail online	Researcher intervention on using device
Multiple thumbnails online	
Multiple thumbnails offline	
Enlarged image (full-screen)	
Photo countdown	
Photo upload (preview)	

This discussion resulted in the creation of the actual list of codes, which is presented in Table 10. This 52 item list of activities (actual codes) included ways the devices and maps were held, different pointing gestures, means of sharing a screen and a device, frequency of stopping or parking for the system use, and effects of a map or phone ownership on a balance of power in each team. At this point the two key researchers watched through the videos again, coding the activities of players observed on the video using the codes created in this phase. The two key researchers had been involved in the planning and implementation of all stages of August 2008 and August 2009 trials so were considerably more involved from the project's inception and particularly at this stage than the rest of the core group.

Table 10. List of activities (codes) searched from the video data.

1. One-handed panning 2. Two-handed panning 3. Short-distance panning 4. Middle-distance panning 5. Long-distance panning 6. Pointing on map 7. Pointing mid-air 8. Marking map 9. Pointing on screen 10. Pointing to environment 11. One-handed hovering 12. Two-handed hovering 13. One-handed selecting 14. Two-handed selecting 15. Checking clue thumbnails 16. Checking thumbnails taken by others 17. Checking enlarged images taken by others 18. Checking own gallery 19. Take photos 20. Checking own locations	21. Sharing the device 22. Sharing the screen 23. Standing 24. Squatting 25. Moving closer to map 26. Moving further from map 27. Use non-Maplens functionalities e.g browsing, sending SMS messages 28. Use offline functionality (multiple photos offline) 29. Use online multiple images 30. Used another phone for other functions 31. Use while walking 32. Use parked 33. Use stopped (things down) 34. Vertical use of device 35. Horizontal use of device 36. Skewed use of device (between H+V) 37. Switch phone/map 38. Switch map/environment	39. Switch phone/environment 40. Interact with researcher 41. Interact with other teams 42. Interact with spectators 43. Vertical map 44. Horizontal map 45. Map on the ground / tabletop 46. Folded map 47. Rolled map 48. Map owner main augmenter 49. Map owner not map augmenter 50. Leadership for phone holder 51. Leadership for map holder 52. Swap roles

While going through the list (coding the video data) researchers marked the frequency of each action by using a 4-point scale:

- Did not occur (not marked),
- Less (L=<3, occurred less than three times)
- Average (X=3-5, occurred 3-5 times)
- More (M=>5, occurred more than 5 times)

In unclear situations questionnaires filled by players were consulted for background information. We then looked through this list to see what commonalities or patterns were emerging between the three conditions: 1/1, solo user; 1/3 shared device; 3/3 device each.

During the interviews and the different passes of the video analysis the core researchers sought new phenomena not already identified. The questions and the list of actions were updated continually, when new phenomenon were identified researchers returned to previous videos and interviews to check and update the findings. This provided us the coding result that was inserted in a spreadsheet, not included here for the purposes of anonymisation.

These results were then used to identify differences between the conditions (for example in “holding maps”). In Figure 7 is a scanned image of our coding process, which how we worked with the data marking the differences between conditions and highlighting important items. This lead us to “drilling down” deeper in the data and counting results, which is discussed in the next subchapter. At this stage we were still looking for patterns or any unusual occurrences of significant difference to emerge. The videos from each team were gone through several times and each occurrence of the activity was marked in the excel sheet.

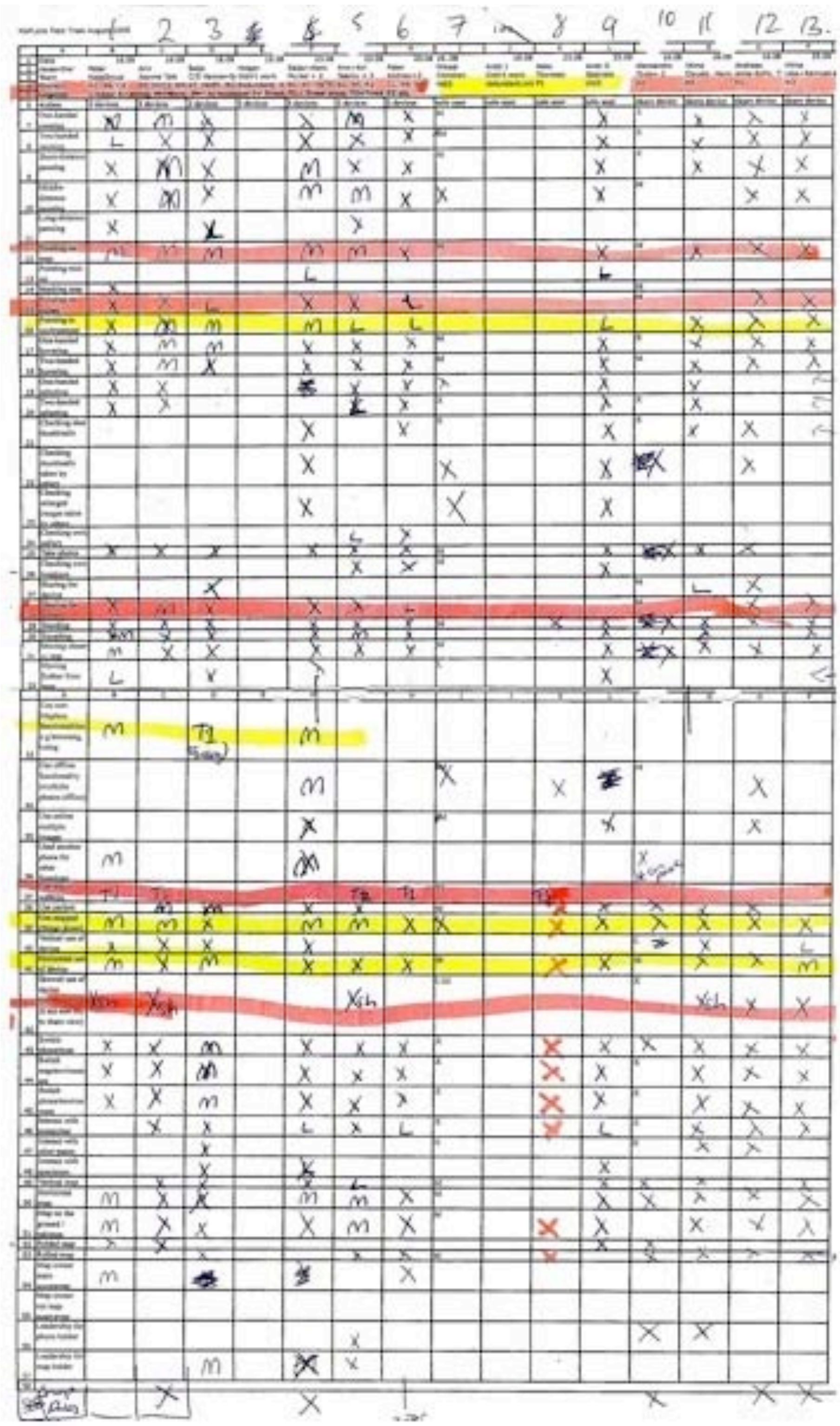


Figure 7. Scanned image of our coding process.

Drilling down

A smaller list relating to specific AR sharing, screen sharing and pointing to screens, map and environment was then compiled and all the footage was gone through to count instances of these activities that are listed in Table 11.

Table 11. List of sharing and pointing related activities

Non map lens user points the map	Phone screen shared with X people

ML user points it's screen	Screen shared: horizontal
Other person points device screen	Screen shared: vertical
ML user points the map	Screen shared: tilted
ML user Quiet when point	Phone moved because of colliding
Other person points the map, while ML used	Phone kept further away to avoid collisions
ML user points environment	Phone used for something else (time)
ML user looks environment	Phone not used (time)
Other person points environment	Changing user (mid.session)
	Using another phone

We decided to put more effort into clarifying the typical ways the device was used during the game, the length of the use sessions and if all the devices were used equally. We also recorded, when the devices were shared between the users, and when they were used simultaneously with other devices, and clarified how this affected how and what they were used and also how that effected the communication around them.

While watching videos :

- We recorded the length of each usage session, where it occurred (indoors/outdoors), and if a map was hold on a hand or put on a bench or ground.
- For each device we recorded/counted:
  - how long it was used during each use session,
  - how long it was used simultaneously with other devices, (see Figure 10)
  - if a phone was moved during a session because it was about to collide with another phone, or if it was purposefully kept further away from other phones ,
  - if it’s screen was shared during the use session
    - how many people it was shared with,
    - if it was kept horizontal, vertical or tilted,
    - how many times device user and other persons pointed it’s screen,
  - how many times device user and other persons pointed to a map to identify or suggest a place of interest ,
    - and if this was done for communication, or to keep the place on mind, when simultaneously trying to identify it by watching around ,
  - how many times device user and other persons looked at environment to identify the place,
  - if a user of the device changed during the usage session,
  - how long a device was used for something else than augmenting (e.g. browsing),
- Were other phones (the personal ones) used during the session and how long

SIMULTANEO	3phones(sec)	2phones(sec)	1phone(sec)		sum(sec)		2or3 phones u	3 phones%	2 phones %	1 phone %
Team A, 1st us	65	55	0		120		%	54.166667	45.833333	0
Team A, 2nd u	34	100	0		134			25.373134	74.626866	0
Team A, SUM	99	155	0	0	254					
Team A, 1st h	38.976378	61.023622	0			team A	100			
Team A, 3rd us	0	34	85		119			0	28.571429	71.428571
Team A, 4th us	0	65	65		130			0	50	50
Team A, 5th us	0	0	40		40			0	0	100
Team A, SUM	0	99	190	0	289					
Team A, 2nd I	0	34.256055	65.743945				34.256055			
Team B	3phones	2phones	1phone		sum(sec)					
Team B, 1st us	99	57	33		189			52.380952	30.15873	17.460317
Team B, 2nd u	0	40	24		64			0	62.5	37.5
team B, SUM,	99	97	57	0	253					
team B, 1st h	39.130435	38.339921	22.529644			team B	77.470356			
Team B, 3rd u	0	24	36		60			0	40	60
Team B, 4th us	0	0	11		11			0	0	100
Team B, 5th us	0	0	113		113			0	0	100
team B, SUM,	0	24	160	0	184			0	13.043478	
team B, 2nd h	0	13.043478	86.956522				13.043478			
Team C	3phones	2phones	1phone		sum(sec)					
1st use	0	0	40		40			0	0	100
2nd use	0	0	75		75			0	0	100
3rd use	0	13	36		49			0	26.530612	73.469388
team C, 1st h	0	13	151	0	164					
team C, 1st h	0	7.9268293	92.073171			team C	7.9268293			
4th use	0	0	30		30			0	0	100
5th use	0	0	42		42			0	0	100
6th use	0	10	29		39			0	25.641026	74.358974
7th use	0	0	5		5			0	0	100
team C, 2nd h	0	10	106	0	116			0	8.6206897	
team C, 2nd h	0	8.6206897	91.37931				8.6206897			
Team D	3phones	2phones	1phone		sum(sec)					
1st use	0	50	280		330			0	15.151515	84.848485
2nd use	14	3	11		28			50	10.714286	39.285714
team D, 1st h	14	53	291	0	358	team D	67			
team D, 1st h	3.9106145	14.804469	81.284916							
3rd use	50	0	0		50			100	0	0
4th use	72	38	0		110			65.454545	34.545455	0
team D, 2nd h	122	38	0	0	160					
team D, 2nd h	76.25	23.75	0				100			
Team E	3phones	2phones	1phone		sum(sec)					
1st use	117	147	138		402			29.104478	36.567164	34.328358
2nd use	0	0	28		28			0	0	100
3rd use	0	130	5		135			0	96.296296	3.7037037
team E, 1st ha	117	277	171	0	565					
team E, 1st ha	20.707965	49.026549	30.265487			team E	69.734513			
4th use	0	0	10		10			0	0	100
5th use	0	0	89		89			0	0	100
6th use	0	0	65		65			0	0	100
team E, 2nd h	0	0	164	0	164			0	0	
team E, 2nd h	0	0	100				0			
Team F	3phones	2phones	1phone		sum(sec)					
1st use	0	0	57		57			0	0	100
2nd use	0	0	18		18			0	0	100
3rd use	0	0	28		28			0	0	100
team F, 1st ha	0	0	103	0	103					
team F, 1st ha	0	0	100			team F	0			
4th use	0	0	53		53			0	0	100
5th use	0	0	55		55			0	0	100
6th use	0	0	35		35			0	0	100
team F, 2nd h	0	0	143	0	143					
team F, 2nd h	0	0	100				0			
Summary	3 phones		2 phones				AVERAGE FC	3 phones	2 phones	1 phone
							1st session	22.608683	21.285124	56.106193
1st half	3 phones, average at beginin		2 phones, average at beginig		sum		2nd session	12.562189	24.640192	62.797619
	17.120899		28.520232		45.64113		3rd session	16.666667	31.899723	51.43361
							4th session	10.909091	14.090909	81.25
2nd half	3 phones, average at end		2 phones, average at end				5th session	0	0	98
	12.708333		12.513865		25.222198					

Figure 8. Counting instances.

In addition to this we identified the typical 'patterns' of use, especially

- o how user's attention switched between the device screen, the map and the environment,
- o how the device, the map and other items were held, and

- how users were aligned during each usage session.

  We compiled the average amount of pointing in the three different team conditions (see Figure 9)

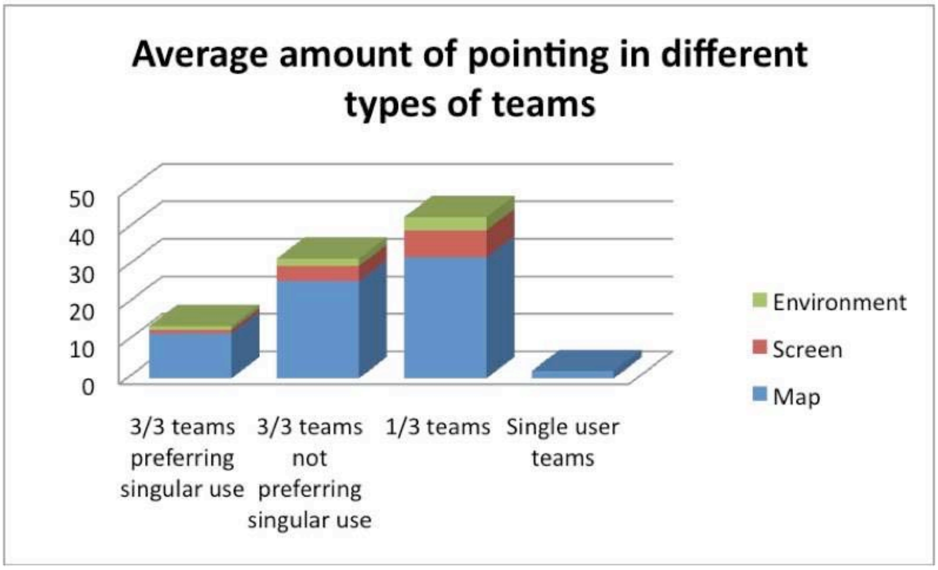


Figure 9. Average amount of pointing per condition

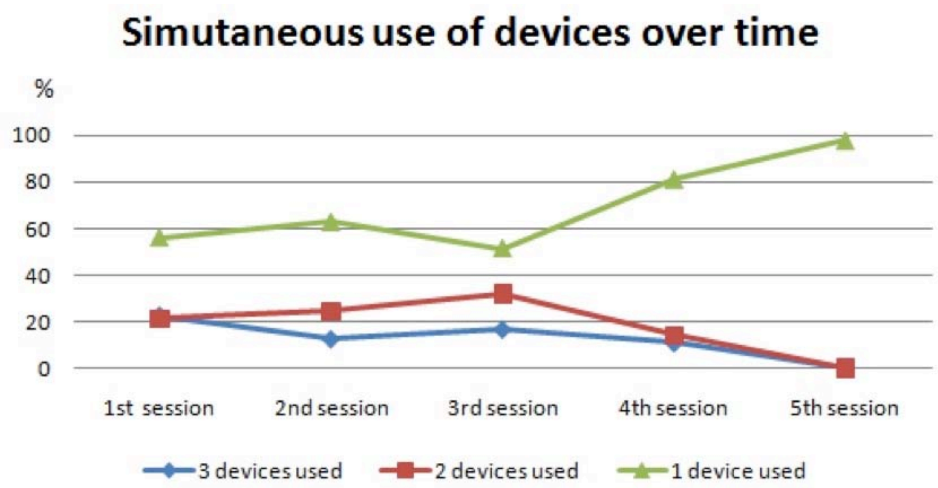


Figure 10. Simultaneous use of devices

We then ascertained the average number of times the main and auxiliary phones were used in 3/3 teams. (see Figure 11)

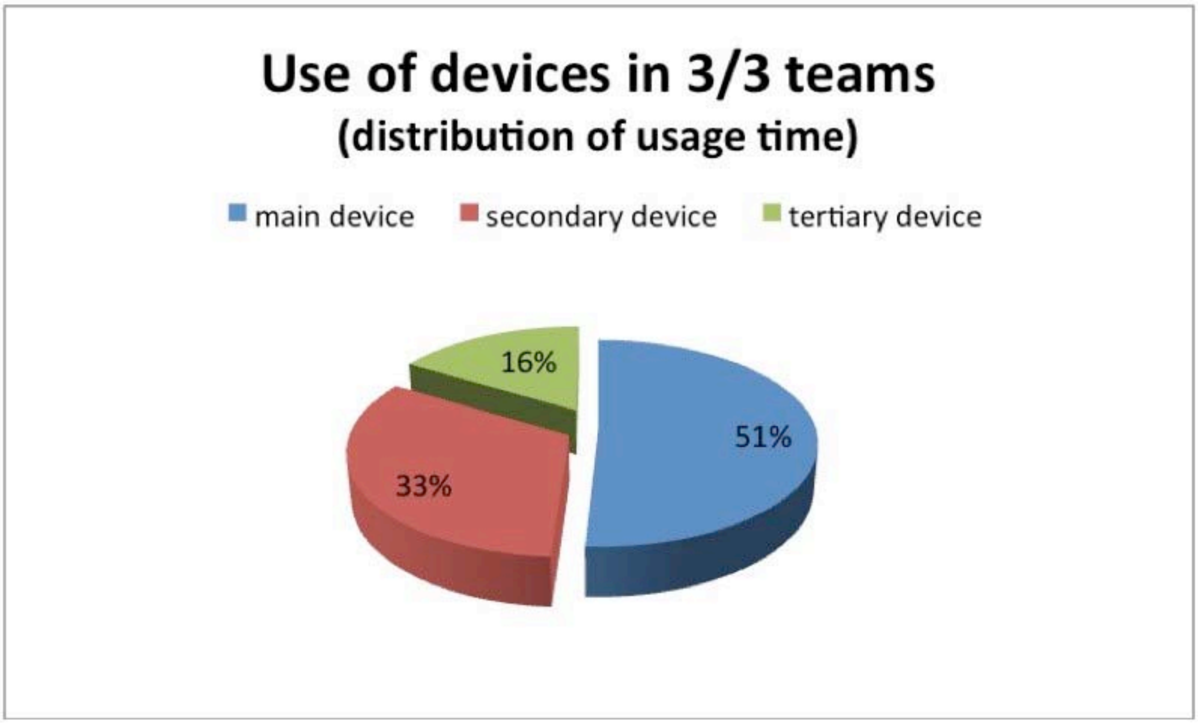


Figure 11. Use of devices in 3/3 teams

Division of Labour and Team Roles

Also, lists for coding of team roles was compiled to capture the essence of how the groups worked and who was in control the group (= who was the alpha user). This coding system can be seen in Table 12, how the division of labour was achieved, what teams were more active (see Figure 12) what configurations of teams spent most time on which activities etc. The researcher marked an instance of each activity for each team (noting there was usually one most active and decision-making (alpha) phone on the map at any one time. The compiled results of these are discussed in the findings section.

Table 12. Coding system for team roles.

Alpha user dominant phone
Auxiliary supporting phone 1
Auxiliary supporting phone 2 / other use (e.g browsing)
Manage map (opens, carries or takes away)
Person who choreographs and makes decision on where to go next in the game
Person who scouts or does other supportive tasks to AR Phone use, for e.g. decides on where to go next.

3 DEVICES		SHARE DEVICES	
R1 Totals	21	P1 Totals	13
R2 totals	23	P2 Totals	12
R3 Totals	13	P3 Totals	26
3 Devices	57	Share Devices	51
K1 Totals	24	A1 Totals	22
K2 Totals	28	A2 Totals	18
K3 Totals	33	A3 Totals	23
3 Devices	82	Share Device	63
M1 Totals	23	C1 Totals	37
M2 Totals	14	C2 Totals	34
M3 Totals	20	C3 Totals	35
3Device	59	Share Device	106
P1 Totals	14		
P2 Totals	9	D1 Totals	5
P3 Totals	18	D2 Totals	3
		D3 Totals	12
3Device	41		
		Share Device	20
N1 Totals	27		
N2 Totals	24		
N3 Totals	25		
3 Devices	72		
Q1 Totals	16		
Q2 Totals	7		
Q3 Totals	11		
3 Devices	35		

Figure 12. Activity levels in each group, to determine dominance or equity in distribution of tasks

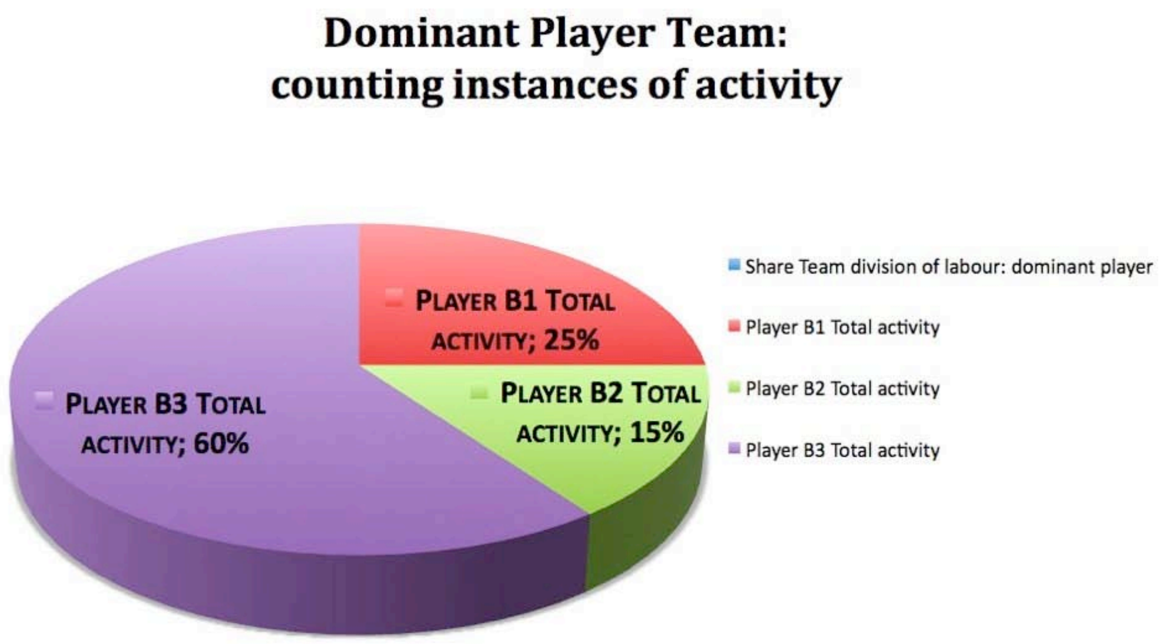


Figure 13. Instances of activity.

Coding MapLens demo use episodes

In the user interviews we had asked the users to show with *MapLens* how they used the system and talk aloud while demonstrating their use. These demo episodes were also analysed by one researcher who coded from the videos the occurrences of how many times users used the old and improved features (see Table 9). We added this information to the spreadsheets with a column for each feature and demographic information such as gender and education level (see the questionnaires in Appendix 1).

3.2.4 System logs

MapLens logging was verbose and produced a lot of information of every system. These logs were then parsed to see how many times certain features were used, as this could not be observed from the video all the time, and to cross check and support findings.

Table 13. Averages of features used by different teams.

Condition	N	Uploads	Enlarged	Enlarged Clues	Enlarged User Photos	Thumbs Viewed Total	Thumbs viewed in normal view	Thumbs viewed in multiple thumbnail view	Entered Multi-Thumbnail View	AR Features used in total
1/1	3	13,33	24,00	6,75	17,25	374,25	269,50	104,75	8,00	460,00
1/3	3	20,33	18,00	8,00	10,00	344,00	275,33	68,67	7,33	389,67
3/3 separately	18	15,39	14,50	4,72	9,78	242,39	202,17	40,22	4,89	277,17
3/3 together	6	35,67	50,17	16,83	29,33	544,50	629,33	260,33	34,17	664,50

Table 13 shows the differences in averages of features used by different teams. Solo users (11) and shared device users (1/3) used the different features approximately the same amount and differences found were not statistically significant.

Although the averages between 1/3 and both 3/3 conditions are different, these differences are not statistically significant. Statistical comparison in general with such small N values is not feasible. For more in depth analysis we would have needed more cases to compare.

Finding no differences in log analysis highlights the importance of the role of qualitative analysis: looking just at the logs we could easily come to the conclusion that there were no differences in use when comparing the different group configurations. But the fact that the users viewed the same amount of thumbnails in general does not tell us much: for example, in which situations did they use the system, how many usage sessions did the users have, what were the differences in roles while using the system? Finding out these kind of differences is only possible with qualitative analysis, observing how the users actually used the system.

3.3 Findings

In this subchapter we review our findings from the analysis process presented in earlier subchapters. Most of these results were written in a publication submitted to CHI2010 conference.

3.3.1 Questionnaires

As shown in Table 14, the data suggest significant differences between single and multiple user conditions (conditions 1/1 vs. 1/3 vs. 3/3). In terms of attention (A2), activity in the environment (A8, A10) and the challenge-skills balance items (B10 to B20) the group configurations (3/3, 1/3) scored higher than single users (1/1). The reported ratings for the enjoyment or loss of self-consciousness (B21, B23) are generally high for all conditions, but group conditions had the highest scores with these items as well. Also, group configurations also reported higher scores in the Intrinsic Motivation Inventory (IMI) part of the questionnaire (C1, C10) than single users.

Differences between the group conditions (1/3 and 3/3) were not as strong as one would expect. The only significant difference between the group conditions was how easy they found the game (B8), three device team members scoring higher than shared device team members.

Table 14. Questionnaire items showing significant differences between the conditions

General Linear Model Pairwise comparisons	Conditions compared	Mean Diff.	Std. Error
A2: The game took most of my attention	1/3 vs. 1/1	2.167*	.992
A8: I felt I could be active in my surrounding environment	1/1 vs. 3/3	1.833*	.843
A10: I thought about whether this map & phone system could be of use to me	1/3 vs. 1/1	3.000*	.998
	3/3 vs. 1/1	2.833*	.962
B8: How to play the game was easy	3/3 vs. 1/3	.833*	.293
B10: I understood how to play the game when I left the meeting room	1/3 vs. 1/1	1.167*	.523
	3/3 vs. 1/1	1.417*	.504
B19: I understood what the immediate tasks were and what I needed to do to achieve them	1/3 vs. 1/1	2.167*	.814
	3/3 vs. 1/1	2.250*	.784
B20: I knew how I was progressing in the game as I was proceeding	1/3 vs. 1/1	2.000*	.643
	3/3 vs. 1/1	2.167*	.619
B21: I was not as aware of time passing or of other people outside of the game as I feel I would usually be	1/3 vs. 1/1	2.667*	.995
B23: I enjoyed putting my feet in the grass, looking at the leaves, testing the pond water and similar tasks	1/3 vs. 1/1	2.333*	.675
	3/3 vs. 1/1	2.417*	.651
C1: I enjoyed doing the game tasks	1/3 vs. 1/1	2.333*	.540
	3/3 vs. 1/1	2.333*	.520
C10: I felt pretty skilled at the game tasks	1/3 vs. 1/1	2.000*	.776

Notes:
*. The mean difference is significant at the .05 level.
All items 1-5 scale; Ax: presence, Bx: Flow; Cx: IMI

3.3.2 Interviews

People reported being very engaged and involved with the game, although for 3/3, half of the users reported the most engaging single thing in the experience was the technology. Almost all users reported having used pointing to the map as a means of communication between team members, and half of them reported pointing being very helpful to refer to items seen through *MapLens*.

3.3.3 Video analysis

Most teams left the briefing room in the museum and completed museum tasks before venturing outside. Some teams were more systematic, planning their route by utilising *MapLens* before heading outside, while others traveled from clue to clue. The first sessions of MapLens use were longer than the later sessions.

MapLens usage sessions, where one or more *MapLens* device(s) was used over one or more map(s), were slightly longer and more frequent in 1/3 teams (mean 1:36 min, 7 times, σ =2.9) than in 3/3 teams (mean 1:26min, 6 times, σ =1.4). The solo users used *MapLens* more frequently, but for shorter sessions (mean 1:06min, 8 times, σ =1.0). 1/3 teams spent a couple of minutes more total time using the technology during the game than 3/3 teams or single user teams.

The use of devices in 3/3 teams

While using the system, 3/3 teams typically deployed just one map, even though they were given three maps upon leaving the museum. Some 3/3 teams began their first sessions by using more than one map (Figure 14c), but soon they adopted the common habit of sharing one map between them.

When starting the game, 3/3 teams used 2 or 3 devices simultaneously, but once familiar with the game and the system, mostly just one device was used. Four of the teams continued to use two or three devices simultaneously over half of the time (mean use time for two phones 33%, three phones 20%) but two teams consistently used only one phone throughout the game (mean use time for two or three phones 4%). Even with the teams that used two or three devices, there was still clearly one ‘main device’ (mean total use 51%) and more ‘secondary or tertiary device/s’ that were used less (mean use 33% and 16%).

As well, two phones seemed to be the maximum amount of devices that could easily fit at one time on the map surface, and here again, one was always a ‘main phone’ with one or two ‘secondary or tertiary phones’. In some 3/3 teams phones collided, so users moved their phones to different heights above the map, moved alongside the other device on the map, or withdrew their device and looked through the other device or sideways under the device with the naked eye to avoid this reoccurring (see Figure 14). To avoid collisions some players also explored different areas of the map (see Figure 14). Also players grouped together to work on the same *MapLens* task, rather than delegating out tasks and working solo. Because of space around the map, and the way teams worked, discussing and solving problem together, we surmise that over time more efficient use emerged with one main phone over the map. We also observed the decreased use of multiple phones over the game’s duration.

Figure 14. (a) 3 devices used simultaneously (fitting better, when on different heights) (b) 2 devices used simultaneously. (c) Using just one device was often more effortless.

Embodied interaction

We refer here to the way in which the game threw players into close physical proximity, forcing them to *cluster* together (draw close together) around the small devices’ screens; they used hands, phones and tangible artifacts to communicate with each other.

Looking

When using MapLens to identify a location, MapLens users in all teams typically switched their attention between the device screen, the map and the environment. Typically, (1) a player first identified a location on a device screen, (2) after that the player checked where the location exactly was on the map, and finally (3) the player looked to the environment to decide where to head next. In addition, all teams also used the shortened version of the method by switching attention just between a device screen and a map, and/or a device screen and the environment. Switching attention between a map and the environment was naturally more typical for non-*MapLens* users than *MapLens* users (see Figure 15).

Figure 15. Attention switching (a-c) from a device to a map and then to environment, (d-f) between a device and environment, (g-i) from a map to environment

Sharing

In 1/3 teams the players shared the device screen practically throughout all sessions (mean total 89%). Players tilted their screen for others to see it better, pushed the device closer to their teammates, handed it over or stood closer together (see Figure **16**). Pointing to the screen used by another team member was more common in 1/3 teams. In 3/3 teams, the intentional sharing of screens happened less, typically only a couple of times during the game, and then only for a few seconds (see Figure **16**).

In 1/3 teams, all three players typically shared the screen almost continuously (79% of the shared cases), while in 3/3 teams, it was most typical that two people shared a screen (71% of the shared cases). Also, pointing to another person's screen was more typical in 1/3 teams than in 3/3 teams (see Figure **16**). Also, 1/3 teams shared information on a map and environment more frequently, while 3/3 teams shared it (whenever they shared) through their own device. All players in the 1/3 teams, both *MapLens* users and non-users, were looking and pointing to the environment more frequently than in 3/3 teams.

Figure 16. (a,b) Sharing a screen and a device in a 1/3 team (c) sharing a screen in 3/3 team

Pointing

While single users pointed to a map typically a couple of times during the game to support their own use and thinking, the *MapLens* users in all multi-user teams pointed to the map average 12 times throughout the game, usually to communicate locations to their team members.

Non-*MapLens* users pointed to a map more in 1/3 teams (mean 14.5 times) than in 3/3 teams (mean 10 times). We surmise that these players were not able to augment the location themselves, and needed to inform or query the *MapLens* user about information. As well, these non-*MapLens* users had their hands and eyes free to look around, whereas people in 3/3 teams were more focused on using and looking through the device. All players in the 1/3 teams were pointing to a map, a screen and an environment more frequently (means respectively 26, 7, 3) than in 3/3 teams (means 22, 3, 2). We observed that for 3/3 teams, pointing on the screen could often be replaced by looking through *MapLens(es)* screens to *e.g.,* augmented information or a finger pointing on a map (see Figure 14a).

Differences in 3/3 teams

The two 3/3 teams, who mainly used just one device for the entire game (at least 92% of the time), also pointed to the map, device screen and the environment notably less (mean 11) than the members of other 3/3 teams (mean 29). The difference was especially clear with the non-*MapLens* users of these teams, who pointed to the map only couple of times (mean 4) compared to the other 3/3 teams, where the mean count was 15. These teams acted differently to the 1/3 device teams.

One 3/3 team used just one device 92% of the total time they used the system. The team had one main augmenter, who knew the city well, but all the team members took turns and used the system actively (see Figure 17a,b). While one team member was using the system, others were browsing, working with a clue booklet and kit, exploring the environment, taking photos and browsing offline the photos taken by other players. Despite clear roles and a singular use of the system, team members often agreed on a next destination by pointing to it on a physical map before moving ahead.

Figure 17. (a,b) 3/3 team members using the system by turns, (c) 3/3 team members preferring simultaneous use

Another 3/3 team used two or three phones simultaneously relatively often (28% and 22% respectively) (see Figure 17c). The team actively discussed and planned their activities and pointed to a map, a device screen and their environment while using the system. This team used two maps simultaneously while inside museum, but switched to one map use when outdoors. They also folded the map outdoors, used *MapLens* parked, and attempted use while walking.

Teamwork

We identified occurrences of predominant recurrent tasks that team players took on. We then compiled this information adding device use time, activity levels and division of tasks.

Division of Labor

We found there were five main tasks, four of which included responsible decision making for the team (agency), and the last broader category included more general support tasks. These tasks were (see Table 15):

(1) *Alpha phone use:* dominant phone used by the team to view and make agreements with

(2) *Secondary or tertiary phone use*: supporting lens in relationship to alpha phone. Support phone roles often switch between the phones, *e. g.*, when one player finds next clue through their phone then this become alpha phone that team acts through;

(3) *Map use:* carrying, orienting, holding. The map begins and ends *MapLens* use; directing game play and time management carrying, orienting, holding, bringing out, or withdrawing the map at any point in the game. Depending on use, this player often had no hands free for *MapLens* use, withdrawing (or offering the map begins and ends *MapLens* use, so this was a major role that directed game play and time management in the team. It needs to be stressed that while players often began with two maps, and in one instance with three maps, after the first initial attempts *all* teams used only one map; (see Figure **20**a)

(4) *Navigation*: decision on where to go next, often several occurrences for each *MapLens* use; route and overall strategy for whole game;

(5) *Scouting*: exploring environment, looking, marking, pointing paper map, clue booklet, discussion etc. Does not occur while holding *MapLens*, map or navigation decision.

Table 15. Team roles; how labour was divided up

People: Looking for team roles—counted instances of:
Alpha dominant phone
Auxiliary supporting phone 1
Auxiliary supporting phone 2 or used other use (e.g browsing)
Manage map (opens, carries and takes away)
Person who choreographs and makes decision on where to go next in the game
Scouting and other tasks that support to AR phone use, map management and decision on where to go next. (Cross-checked also with A: Field Trial Video Analysis: Actions)

We found in 3/3 teams, the average for division of labour for instances of use was: 37% phones (with 20% alpha phone, 7% auxiliary 2nd phone, 6% auxiliary 3rd phone); 27% managing map, 26% scouting and other tasks and 14% deciding where to go next. The average amount of time all phones were used across all 3/3 teams was 920 seconds.

We found in 1/3 teams, the average for division of labour for instances of use was: 36% scouting and other tasks, 27% first phone use, 24% managing map and 13% deciding where to go next. The average amount of time phones were used across all share-device teams was 647 seconds.

In descending order of activity for 3/3 teams, we found values for phone use, map use, scouting and navigating (37%, 27%, 26%, 14% respectively). The average time for all phone use for all 3/3 teams was 15.3 minutes (920 sec.). For 1/3 teams, we found in descending order of activity similar values for scouting, phone use, map use and navigating (36%, 27%, 24%, 14% respectively).The mean time for all phones for all 1/3 teams was 10.7 minutes (647 sec.)

Therefore, we found our 3/3 teams averaged 10% more phone use, and 4.5 minutes more phone time compared to the 1/3 teams, who engaged in 10% more scouting activities and other tasks (including pointing and looking to map and environment etc.), and 3 % more map use.

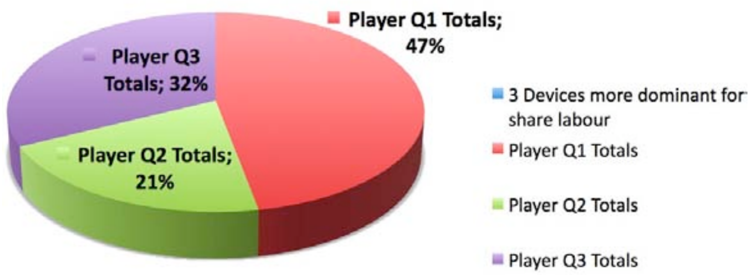


Figure 18. Case example: activity counts 3/3 team with dominant player taking up 47% of the activity

Types of teams

We then averaged the number of occurrences of activity per player per 3/3 and 1/3 team, to gauge how equally players participated in the tasks. After this we categorised the teams according to what these figures revealed. While watching the video we had discussed teams' experiences, in particular noting those teams who easily shared tasks. Below we add these observations to the figures.

Both the 3/3 and 1/3 teams fitted three pattern types:

(1) *Agile*: equal counts of activity. We also observed in these teams that roles flowed from one to the other almost seamlessly (see Figure 17).

(2) *2-share predominant*: two players had larger activity counts, and one player less active in the game. however all players were relatively active *e.g.,* Player 1=23, Player 2=22, Player 3=13. However, all switched roles and took turns at tasks in the spirit of team play. (see Figure 19

(3) *Controller*: one player with much higher activity counts than other two. (e.g., Player 1= 26, Player 2= 11, Player 3= 7). Roles were often fixed from game start, with the controller being reluctant to share agency **(**see Figure **20**b,c)

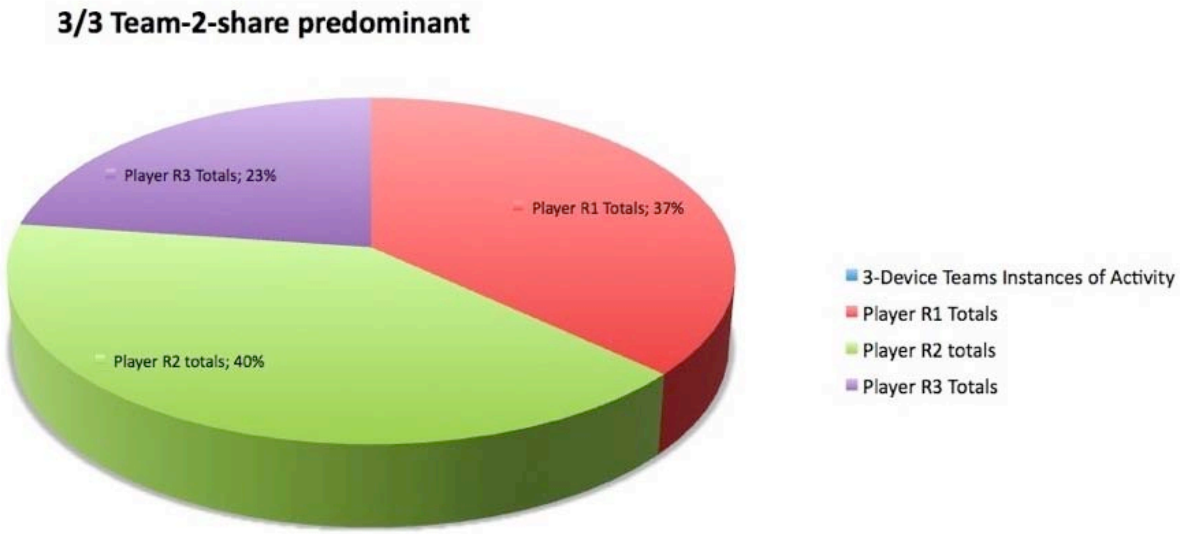


Figure 19. Shared labour in 2-share predominant team

We found that the 3/3 teams consisted of two *agile*, two *2-share predominant* and one *controller* types. In 1/3 teams, there was one *agile*, two *2-share predominant* and one *controller* type. It appears that team type was not determined by the number of devices, but rather by the personality make-up of teams. In the two *controller* teams, the dominant player often put the device back in the pocket, or while using it, hid the screen from team view. We rule out shading from direct sun, as we saw no other instances of this kind of use in other teams (**Figure 20**b). Obviously a *controller* in a 1/3 team impacts the general team experience more heavily than in a 3/3 team, where the other devices can be used.

Figure 20. (a) She is still using map but the map carrier pulls map away and team moves on. (b) Controller: he even hides the screen view from his team, (c) and puts device in pocket so he is always in control.

General observations on teamwork

The map was as essential as the device (the pair needs to work in tandem), so as with the *MapLens* in lead use, the holder of the map (or device) could choreograph game play (when and where they moved next) by sweeping the map in or away (**Figure 20**a). In the two *controller* teams, we surmise either players did not intervene as they were too polite, happy to take a lesser role or unfamiliar with outdoor use of the device. In *2-share predominant* teams, the predominant two players either knew each other beforehand or connected while playing the game, but also made sure they included the third person. In *agile* teams, players did not necessarily know each other beforehand, but managed the sharing of tasks in an equitable manner.

Impact of new AR features on use

Compared to the previous version (Morrison et al., 2009), technical improvements in *MapLens2* provided users with more flexibility of use and a larger range of functionalities, which were used extensively during the game. The use of the new features meant:

Almost every user browsed the map for clue images (94% 3/3, 100% 1/3, 75% 1/1), and many browsed for their own and other team's photos (47% 3/3, 70% 1/3 50% 1/1), even though they considered it a side-activity not connected to the game. Most users browsed photos on a scrollable overlay view developed for the system to support the viewing of dense image areas (70% 3/3, 70% 1/3 and 50% 1/1). Some players also browsed photos while walking, because this was now possible without constantly needing to access the map. The device over the map allowed defining an interest area, connection to the content server to download, and 'caching' all images within the selected area, which were then available for offline non-AR browsing. This version of '*Take-away Interface'* was perceived as most useful by those who employed this technique. Most users also found the "you are here" icon showing a live GPS position of their position helpful (70% 3/3, 58% 1/3 and 75% 1/1).

Use outdoors

The teams either used the system while standing and holding a map, or they put a map on a supporting surface (see **Figure 21**). When comparing *MapLens2* with *MapLens1* (Morrison et al., 2009), users no longer needed to stabilise either the map or their hands when augmenting. The system also worked well for a wide variety of angles distances between map and device. While there was increased capacity to use the device while walking, we found only four 3/3 teams, and two 1/1 teams attempted to do this, but did not continue use while walking outside the museum.

Figure 21. a,b) Different ways to use MapLens: players either stand and hold a map or lay a map on a supporting surface like ground, table or a bench c) using while walking d) a team struggling with 3 maps at the beginning of the game.

Outside use with wind, direct sunlight and obstacles, such as people passing, uneven pavements, moving traffic, as well as wind and direct sunlight impacted on *MapLens* use. We observed constant negotiation with wind and large maps. Despite the fact that the map was an inexpensive item and could obviously be stressed without consequences, only two 3/3, one 1/3, and one 1/1 team folded the map, making it easier to use.

More temporary stopping and more mobility

Just as 'no parking' zones are used for temporary parking, all our players exhibited more 'park type' activity (stopping briefly to check a detail) and continued to stand while using *MapLens* (see Figure **16**c, Figure 15d-f). Two of the 1/3 and one solo users only used *MapLens* while standing. All teams also made longer stops, using the device for extended periods and putting down items (see Figure 17a-c, **Figure 21**b). However, the recurring temporary stopping for quickly checking a detail was a phenomenon not found in our last trial, where more *place-making* was the common practice. We believe the ability for quick stops is a result of the technical improvements of our system.

3.3.4 Conclusions

Trickle Down Effect of Robustness

We return now to the central tenet of our previous study, where we found teams positioned themselves in close bodily proximity around the device and map (a phenomenon we identify in this paper as *clustering)*, in order to render AR information visible to all, and so enabling collaboration. A key question left unanswered by MapLens1 was how does group use change if technology is more flexible and users have multiple mobile AR devices? For MapLens2, we made a deep investigation to ensure place-making and collaboration are not just the by-product of poorly working technology requiring people collaborate to get results. MapLens2 uses a revised tracking technology combining detection, ADDING incremental tracking, and yielding a much more robust and usable result. Who uses the phones and how (the nature of the collaboration) depends on team composition and personalities. With *MapLens1*, we saw team personalities impacted use but had no evidence, so for *MapLens2* we ensured we could prove impact, adding to an understanding of situated collaboration around mobile devices.

With *MapLens2,* our user-centered design improvements from *MapLens1* trials included improvements in stability and range, hugely improving the team's degrees of freedom in positioning together. Users no longer needed to *cluster* so closely to the map, or to stay still. With *MapLens2*, we see that a more robust system facilitates more temporary stops ("parking"), standing while using, and that use from a distance is made possible. The design of improved AR factors can positively affect team configuration and usage patterns.

Four conclusions can be drawn on the basis of our study:

1. *Solo use was not the main method employed*. Given the opportunity to establish common ground through shared space, teams will be compelled to do so. AR applications should aim to simplify place-making as much as possible. As a result for future use of this game,

for 5 teams of 3 players, we surmise we can dispense 8-10 phones for optimal use, where e.g.15 phones is over-kill.

2. *Multiple devices are useful but not essential for collaboration*. But they extend the range of interactions and thereby give much better opportunity to collaborate and potentially overcome problems such as overly dominant users. It is therefore worthwhile supporting multiple AR devices. Multiple devices expand the ways that groups can collaborate (additional tasks, asynchronous, not co-located) and how phones are used. We do not see this as a contradiction, rather as an expansion of potential collaboration and place-making styles
3. *Agency enabler*. Personal AR device increases enjoyment of the use and of the collaboration. This has very likely a strong positive influence on the long-term motivation of using AR on mobile phones.
4. *Take-Away Interface,* where users can "steal/ cache" AR information from the source and then walk away and access this information while off-line. This was found to be an exciting and useful development. This also removes the need to constantly use the device in tandem with tangible maps or other information artefacts, and reflects the more lightweight pattern of use: "park/investigate", "move/investigate further", that also lessens cognitive load. Being able to see AR information while off-line is a major step and novel use for AR. To be able to use Take-Away AR in combination with hastily stopping can impact ease of use with little disruption to everyday activity.

It does appear that teams were less inclined to be experimental due to time concerns imposed by game play. For example they did not stop and fold the map into different configurations to lessen wind impact and many did not try again while walking or see where other players were with off-line browsing. A less-pressured scenario may have generated more use of all AR features.

Multi-lens and Single-Use: how the teams AR'd

Five main conclusions can be drawn:

1. AR on mobile phones is easily used in multi-user situations. Multiuser teamwork provides a more satisfactory experience than solo use.
2. A single and shared physical frame is preferred for teamwork over multiple frames.
3. Collaboration in the multi-lens situation is characterised by sharing of AR information among the members through displays that are visible to others. This decreases the amount of communication work needed to reach consensus.
4. Even when multiple lenses are simultaneously used, one lens emerges as dominant.
5. In comparison to the previous year's study with *MapLens1*, *MapLens2* succeeded in supporting more flexibility and mobility for individual users and teams.

First, the general finding is the confirmation and extension of our previous result that AR on mobile phones is a natural platform for collaboration. The presence and experience questionnaires show systematically lower scores in comparison to the multi-user situations. Solo users reported enjoying the game less than the multi-user teams. We saw social presence and 'as if real' activity in situated use of AR phones. Technical improvements enables this as MapLens1 tracking instability did not allow this much freedom.

Second, multi-frame use: We found that the greater freedom of use provided does not mean teams will not collaborate. Despite the availability of multiple maps, multi-lens teams still shared the one map and gathered around it. They brought their devices, pens, fingers and clue books—the ingredients for common ground interaction—for pointing and display; their ideas and strategies for discussion and negotiation. We found that all teams discarded their surplus maps and shared just one map, after at most brief use with multiple maps outdoors. Usually, AR is designed to display content relative to a particular unique object. In previous work on collocated AR, users were normally constrained to a canonical shared space, *e. g.*, augmentation of a historical artefact, so the question of whether or not to share space did not pose itself. Our observations show that in practical situations, the option of having multiple individual frames of reference is not relevant. Users, in particular those who *cluster* together for collaboration prefer a single physical frame of reference. Normally, AR displays content relative to a unique known object e.g., a church in the environment, so the option of multiple versions of that object does not present. However, in offering multiple maps, we offered the opportunity to work with multiple frames of reference. For efficient work-flow, as there was interaction work load involved with using the maps, it made more sense for a team to share the one map (and the workload of maintaining the map in the condition ready for use while on the move) in the game set up. However, there are scenarios where multiple frames of reference would be relevant. For example in remote collaboration with AR, for each party to have their individual frame of reference as well as shared one, would make work-flow easier. Offering multiple maps in a co-located space revealed a preference to establish a shared single-frame space. We feel this is an important finding for general understanding of AR reference frames.

Third, we witnessed these teams creating a common platform for co-operation (regardless there was not one designed for them). They shared differently than the single-lens teams, as they shared more *through* the device screen, and consequently needed to point less (to the map, the screen and the environment). This was the more efficient way of sharing with the most information at hand, and in view. Having more phones clearly allowed for more variety in use of the devices (including web browsing and phone calls), as well as different styles of interaction and collaboration. Two phones seemed to be the maximum amount of devices easily fitting simultaneously on a map, additional phones were frequently moved up or to the side. A situation requiring users to operate three phones on different parts of the map did not happen, and is unlikely as it contradicts the desire to share tasks. The common situation then involved three users around one map, all working on solving one clue simultaneously. Collaboration emerged spontaneously, and people did not separate out tasks. It seems for multi-lens use, sharing a screen and holding one's own personal device heightened efficiency in collaboration. With 3/3 teams using a singular lens, the communication level was much lower, *e.g.,* in these teams pointing occurred less than half of the time compared to multi-lens 3/3 teams, and less than one third of the time compared to singular lens 1/3 teams. Thus, personal device was literally *at hand* –to put it away, so as to not use it would disrupt natural flow. To use one's own phone also avoids getting in each other's way physically. These were considerations we noted with the evolution of a multi-lens common ground system of use. The *MapLens* system (device in tandem with map) was the most often used artefact and information aspect for the establishment of common ground between the players. In the case of *MapLens*, this type of utilisation seems to have had a peak value at "2 phones". However, we speculate this peak is dependent on the size of the shared space, the distance to the augmented object (consider augmenting the façade of a historical building) and the information density in this space. Pointing directly to the map or screen while using one or multiple phones demonstrated a certain degree of "perceived non-mediation" even on a tiny phone; users pointed to an object through the phone and others looked at what was being pointed at, not at the pointing finger. We saw social presence and 'as if real' activity in situated use of AR phones.

Fourth, regardless of the number of phones available to a team, there was one dominant phone that facilitated the viewing and the decisions. While this phone could change depending on *e.g.,* through which device the latest clue was found, it was consistently found that there was one *alpha* phone on the map surface that all other phones acquiesced to at any one given time.

Fifth, the improved stability of *MapLens2* means we maximise experience and engagement by increasing mobility in two ways. First with enabling more 'parking' or transient use and second, a take away interface also increases mobility, so users can be more active on the move. Both these methods extend access to ad-hoc collaboration, and while we referred in our last study to rapid placing-making, here we have added options with transit stops ('parking'). We extend the discussion on place-making to include a finer grain analysis and identify more agile and mobile forms of place-making more Agile, as place-making is no longer ESSENTIAL just to get the system working (Benford et al, 2005). For more strategic game play or for longer consultation processes, place-making works best. The designed improved AR enhancements extend the scope of available collaborative practices with AR on mobile devices.

Summary and future work

The user study we report here opens up a new domain of investigation, because it extends the study of collaboration on mobile phone based AR to a multi-AR and multi-FoR (frame of reference) situation. Having multiple phones and multiple maps on a team created a novel situation not studied in context of AR before. Normally, AR displays content on a unique object that can be shared by a group. Today's mobile phones allow a wider approach. Not only can each user have her own frame of reference (e.g., the world or the map), and the lens, but there will also be as many FoRs and lenses available in the proximity as there are users simultaneously using the system.

The results of the study lend evidence for the claim that, despite their small physical size and being regarded as personal devices, mobile phones are preferred and voluntarily adopted as means for collaborative efforts in small groups. What we find is that having multiple FoRs is not necessary, but having multiple lenses can be useful. While utilisation of multiple lenses is not linear with the number of users (*i. e.*, not all users use their phones equally all the time), the quantity and quality of collaboration is changed in a number of ways. It is therefore worthwhile to leverage the ubiquity of phones for more collaborative interaction design.

To generalise these findings to real-world use, as we progress to 'serious' outdoor AR use, we will note subtle but important changes in what is used, and how it is used. From the perspective of serious application development, the AR design space is as yet hardly known, so these seemingly smaller findings can be important. From the perspective of interface design and interaction design, neither can we yet know what types of scenarios of use, we are likely designing for. In further work we could compare these findings more explicitly to other AR tasks or experiences, such as e.g, use within a non-AR game (playing the same game but just with a paper map, clue book and kit), using maps/ frames of references of various sizes (see Rohs, 2009 for a study using maps of various sizes), and to restrict use to tasks that require only solo, non-collaborative work to more richly understand and interrogate these findings.

This study also suggests a number of further questions to be examined. One such question concerns the relationship of using multiple devices with the size and structure of the shared space. Another important question that was not considered yet in this work is how showing different, personalised augmentation content on the individual phones – called *subjective views* in the literature – affects the simultaneous use. This is particularly interesting if users are able to create or manipulate virtual content in the environment, which was not part of the experience in our study except for photo taking. The very positive results of spontaneous, voluntary, expanded and lightweight place-making and intense collaboration suggest that there should be more interaction design that evoke these properties, with or possibly without AR interfaces.

4 Conclusions

In this document we have described the general analysis process for evaluating Mixed reality applications and presented an example evaluation case of the *MapLens* field trials. The user experience of Mixed reality technologies consists of many parts including the technology itself, the physical surroundings it is used, the people it is used with and the experience created when the technology is used to mix the physical world with virtual elements. Therefore we cannot rely only on one specific method for evaluating this experience–multiple methods triangulated together are also needed.

Our *MapLens* trials have revealed us that video analysis can be a valuable tool for obtaining information from real life like use of mobile AR. It allows drilling down to even the smallest nuances of the use of the technology including the social and physical aspects of the environment, revealing phenomena, which might have not been otherwise found. In our case the questionnaires were not found to be that useful when evaluating the technology, as the answers reflected more the whole experience than use of the technology.

Looking back at our analysis process, next time we would probably do things a bit differently: the heavy use of video cameras is not always a good thing as it forces the researchers to observe everything “behind the lens”, narrowing the scope of perception for them. For our next trials, the ratio of video cameras and still cameras could be 60% video/40% still cameras, which would allow the core researchers to observe the teams in a more holistic way without having to hassle with the video cameras.

For future work and next trials we would also do more rigorous usability testing before the trials. Finding out limitations in the UI or the system crashing because of heavy load should be found and resolved well before the field trials. Arranging large trials with teams of participants, researchers, phones, arranging museum open, game tasks etc. is complex enough without the technology failing to work at the critical times. This is crucial otherwise this lengthy ground work is wasted, as trials are often temporal events, so organising teams of people with approx. correct demographic balance, booking museum, permission to film with café and internet passes, buying food etc cannot just be postponed and reinstated on short notice. The logistics escalate with the complexity of the study and the number of users and locations etc involved.

We propose *a three-step approach*, customisable on a case-by case basis. We propose that by integrating these three stages into the evaluation process/es, we can progressively implement iterative changes more thoroughly with technology use over time, in real conditions and with multiple users. As a result for future trials we will adopt this three-step approach, which will include:

1) A series of iterative informal lab testing of devices with set tasks that replicate game tasks with a small set of participants. Correction of perceived faults. Small test trial run synchronously on several devices at the one time (e.g. with 5 participants and 5 devices simultaneously). Makes sure that the technology is usable in general.

2) Usability study in outdoor conditions with simple tasks that replicate real use, stressing the use with a worse-case scenario. For example, if the field trial will have 20 participants using the system at the one time, then 20 researchers should use the system at the same time with tasks that replicate game functions. Further, if we expect that 20 phones will take 10–50 photos each during the trial, then set this taking number of photos as a task (and stress the condition, i.e. use worse case scenario where all players take photos in a 10 minute period—as this also tests the upload, GPS locating upload etc.). Other tasks for *MapLens* use would include browsing for locations, looking at photos, looking at multiple photos (online and offline), trying to use map in windy conditions (choose a windy corner for one task), browsing on web browser etc) and for the same set period of time as the trial will take place. Essentially the aim is to force the circumstances so that any breaks in the system occur now. If the system fails, then time is taken to fix the problems and again the system and how it is tested in this stage

2 is replicated until there are no breakage problems. We estimate that depending on the complexity, this stage adds at least two to four weeks to the implementation cycle (depends on complexity and robustness of the prototype).

3) Field trials (as done) with amendments to percentage ratios of video/ still cameras. As well we would look to do a shorter game that tests and expands the capacities of the technology more. The last game was designed to compensate for initial unstable prototypes. As the prototype has become more robust, we can now include and more accurately directly test specific aspects of the technology.

The three step approach for evaluation process is summarised in Table 16.

Table 16. Three Step Approach for evaluating technology prototypes.

Three Step Approach to Evaluation Process	
Informal Lab test	Check that things are working
Usability test to check robustness	Task orientated approach with single aspects checked for robustness at one time (emulate real world use for each aspect, one thing at a time—include e.g., battery life, number of users at one time etc.).
Field test to emulate real world use	Multi-tasking with multiple distractions and activities occurring at the same time. Impossible to relate task to outcome, and care taken to ensure use happens while distracted or among other activities including e.g., social interaction.

References

Benyon, D., Smyth, M., O'Neill, S., McCall, R. and Caroll, F. (2006). The Place Probe: Exploring a Sense of Place in Real and Virtual Environments. *Presence*, 15, 6, 668–687.

Costello, B., & Edmonds, E. (2007). *A study in play, pleasure and interaction design*. In Proc. Designing pleasurable products and interfaces.

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. Harper and Row, New York.

Deci, E. L., Eghrari, H., Patrick, B. C., & Leone, D. (1994). Facilitating internalization: The self-determination theory perspective. *Journal of Personality*, 62, 119–142.

Deci, E. L., Ryan, R. M. The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. (2000). *Psychological Inquiry* 11, 227–268. Questionnaires: http://www.psych.rochester.edu/SDT/measures/IMI_scales.php

Dourish, P. (2001) *Where the Action is: the Foundations of Embodied Interaction*. London, MIT Press.

Erickson, F. (1992). The interface between ethnography and microanalysis. In LeCompte, M.D., Millroy, W.L. and Preissle, J. (Eds.). *The handbook of qualitative research in education*. Academic Press, San Diego, CA, 201–225.

Glaser, B.. & Strauss, A.. (1967). *The discovery of grounded theory: strategies for qualitative research*. New York (N.Y.), Aldine de Gruyter.

Gutwin, C. and Greenberg, S. (2002). A Descriptive Framework of Workspace Awareness for Real-Time Groupware. *Computer Supported Cooperative Work*, 11, 3, 411–446.

IPerg (2005). The official IPerG (Integrated Project on Pervasive Gaming) web site available online: http://iperg.sics.se/

Jackson, S.A., and Marsh, H.W. (1996). Development and validation of a scale to measure optimal experience: The Flow State Scale. *Journal of Sport and Exercise Psychology*, 18, 17–35.

Jegers, K. Pervasive game flow: understanding player enjoyment in pervasive gaming. Proc CIE 5, 1 ACM Press (2007), Article 9.

Jewitt, C. (2009). What is Multimodality? In Jewitt, C., et al. (Eds) *The Routledge Handbook of Multimodal Analysis.* Routledge.

Klopfer, E. (2008). *Augmented Learning Research and Design of Mobile Educational Games*. MIT Press, Cambridge, MA, USA.

Laurier, E., Strebel, I., and Brown, B. (2008). Video Analysis: Lessons from Professional Video Editing Practice. *Forum: Qualitative Social Research* 9, 3, Art. 37.

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer Mediated-Communication*, 3 (2). Available online: http://www.ascusc.org/jcmc/vol3/issue2/lombard.html

Merleau-Ponty, M. *The visible and the invisible*. Northwestern University Press, Illinois, USA, 1968.

Morrison, A., Jacucci, G., Peltonen, P. (2008). *CityWall: Limitations of a Multi-Touch Environment*. Public and Private Displays workshop (PPD'08), AVI 2008: the International Working Conference on Advanced Visual Interfaces. Naples, May, 2008.

Morrison, A., Peltonen, P., Kleimola, J., Jacucci, G., Samperio, R., Saarikko, P., Juustila, A., Reitmayr, A. (2008). *First Environmental Awareness prototype applications.* Deliverable D7.3, IPCity Internal Document.

Morrison, A., Oulasvirta, A., Peltonen, P., Lemmela, S., Jacucci, G., Regenbrecht, H. and Juustila, A. (2009). *Like bees around the hive: a comparative study of a mobile augmented reality map*. In Proceedings of the 27th international Conference on Human Factors in Computing Systems (CHI '09) pp. 1889–1898.

Morrison, A., Lemmela, S., Oulasvirta, Schmalstieg, D., Peltonen, P., Mulloni, A., Regenbrecht, H., Jacucci, G. and Juustila, A. (2009b). *From Single to Multi-Lens Collaborative Augmented Reality on Mobile Phones*. Submitted to CHI2010.

Oulasvirta, A., Jacucci, G. and Saari T. (2006). *Presence Study Manual for IPCity.* IPcity Internal Document.

Peltonen, P., Kurvinen, E., Salovaara, A., Jacucci, G., Ilmonen, T., Evans, J., Oulasvirta, A., and Saarikko, P. (2008). *"It's mine, don't touch": Interactions at a large multitouch display in a city Center*. In Proc. of the SIGCHI conference on human factors in computing systems (CHI'08), ACM Press, New York, pp. 1285–1294.

Pink, S. (2007). *Doing visual ethnography*. Sage, London.

Robertson, T. Cooperative Work and Lived Cognition: A Taxonomy of Embodied Actions. In Proc. ECSCW 1997, K Kluwer Academic Publishers (1997), Dordrecht, 205–220.

Rohs, M., Schöning, J., Schleicher, R., Essl, G., Naumann, A., Krüger, A. (2009). Impact of Item Density on Magic Lens Interactions. 10th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI), Bonn, Germany, September 15-18, 2009.

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, 43, 450-461.

Slater, M., and Steed, A (2000). A virtual Presence counter. *Presence: Teleoperators and Virtual Environments*, 9, 413-434.

Slater, M. and Garau, M. (2007) The Use of Questionnaire Data in Presence Studies : Do Not Seriously Likert . *Presence: Teleoperators and Virtual Environments*, 16, 4, 447–456

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks, CA: Sage.

Sweetser, P., Wyeth, P. Gameflow: a model for evaluat-ing player enjoyment in games. Proc. CIE 5, ACM Press (2005), 1–24.

Tang, A., Tory, M., Po, B., Neumann, P. and Carpendale, S. 2006. Collaborative coupling over tabletop displays. In Proc. CHI 2006, ACM Press (2006), New York, 1181–1190.

Vorderer, P., Wirth, W., Gouveia, F. R., Biocca, F., Saari, T., Jancke, F., et al. MEC-SPQ: Report to the European Community, Project Presence: MEC (IST-2001-37661), 2004.

Wagner, I., Basile, M., Broll, W., Gammon, Jacucci, G., Kuutti, K., McCall, R., Morrison, A., Schmalstieg, D., Terrin, J.-J. (2009). *Consolidated approach to studying presence and interaction*. Deliverable D3.4, IPCity Internal Document.

Wagner, I. (2009) *Guidelines for the set up and analysis of field trials*. IPCity Internal Document.

Appendix 1: Questionnaires for MapLens Users

	IPCity **Integrated Project on Interaction and Presence in Urban Environments** ipcity.eu

Demographic Questions:

Age Groups: which of the following categories includes your age?

<18 ☐	18 – 24 ☐	25 – 34 ☐	35 – 44 ☐	45 – 54 ☐	55 – 64 ☐

Gender: Female ☐ Male ☐

Education Level: ____

Current Professional Status (e.g student, researcher etc): ________________

IT knowledge:

Basic ☐ Average ☐ Expert/Work in Industry or related field ☐

How many hours do you spend with technology each week? (e.g computer, phones, video games, digital camera etc): ______________

Do you know Helsinki well? Yes ☐ No ☐

Are you knowledgable/concerned about environmental issues? Yes ☐ No ☐

Can you navigate easily usually? Yes ☐ No ☐

Do you use a mobile phone regularly? Yes ☐ No ☐

If yes, how many different models have you used?______________________

Any brand most used? e.g. Nokia, Motorola ______________________

What do you use your mobile for? (e.g MMS, Web browsing, games, music)

_______________ __________________________________

Do you use internet/phone/both/maps to find info on a regular basis? Yes ☐ No ☐

If yes, more than 3X per week? Yes ☐ No ☐

Do you use digital maps, e.g. Google Maps? Yes ☐ No ☐

If yes, how often? ______________________________________

Do you use GPS? Yes ☐ No ☐ If yes, how often? __________________

I give permission for images and/or videos to be taken of me during the evaluation

Yes ☐ No ☐, and to be re-produced and used for research publications: Yes ☐ No ☐

For each of the following statements, please indicate how true it is for you, by marking "X" in the box that best expresses your feelings using this scale:

1: Totally disagree **2**: I disagree **3**: I neither agree nor disagree **4**: I agree **5**: I totally agree

1 I concentrated on the map & phone system Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
2. The game took most of my attention Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
3. I understood the over-layed map through the map & phone system Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
4. I could understand how the map & phone system related to the real world environment Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
5. The 'you are here' helped give me a better reading of where I was in the environment Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
6. The map & phone system helped me understand where other teams were located Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
7. I felt as if I was in the same game space as other teams when using the map & phone system Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
8. I felt I could be active in my surrounding environment (move, use the mobile phone and switch from task to task) Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
9. Using this map & phone system made me think in new ways Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
10. I thought about whether this map & phone system could be of use to me Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
11. I didn't really pay attention to errors in the map & phone system Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
12. I am generally interested in new types of digital systems Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
13. When someone shows me a map I am able to imagine the space easily Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
14. It's easy for me to imagine a space in my mind without actually being there Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
15. When someone describes a space to me, it's usually very easy for me to imagine it clearly Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree

For each of the following statements, please mark "X" in the box that best shows your response. Use the same 1-5 scale as shown earlier:

1: Totally disagree **2**: I disagree **3**: I neither agree nor disagree **4**: I agree **5**: I totally agree

1. I was focused on the tasks throughout the game Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
2. I lost focus on the game at times Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
3. The map & phone system quickly took my attention Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
4. The game quickly took my attention Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
5. There were unnecessary game tasks Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
6. Using the map & phone system got easier with time Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
7. Some game tasks were very difficult to solve Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
8. How to play the game was easy Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
9. How to work the map & phone system was easy Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
10. I understood how to play the game when I left the meeting room Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
11. I enjoyed learning how to play the game Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
12. I enjoyed learning how to use the map & phone system Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
13. My skill with the map & phone system increased as I progressed Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
14. I felt I/we could work out my/our own way to play the game Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
15. I could start, stop, or go back to earlier tasks easily Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
16. Errors with the map & phone system did not affect my progress in the game, and could be easily corrected Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
17. I understood the overall goals of the game Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
18. I understood how the map & phone system worked Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
19. I understood what the immediate tasks were and what I needed to do to achieve them Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
20. I knew how I was progressing in the game as I was proceeding Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
21. I was not as aware of time passing or of other people outside of the game as I feel I would usually be Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
22. I became enthusiastic about the game, e.g. wanting to get there first, see the results of the water tests etc Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree
23. I enjoyed putting my feet in the grass, looking at the leaves, testing the pond water and similar tasks

Totally Disagree	☐	☐	☐	☐	☐	Totally Agree

24. I chatted and made suggestions and helped out my team members
Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree

25. I helped other players in other groups Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree

26. I/We actively competed with the other teams
Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree

27. I was aware of the other teams in the environment and that we were all part of the same game
Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree

How did this awareness (or not) of other teams, and all being part of the same game feel to you? Mark "X" to the scale that best expresses your feelings

Impersonal	☐	☐	☐	☐	☐	Personal
Cold	☐	☐	☐	☐	☐	Warm
Ugly	☐	☐	☐	☐	☐	Beautiful
Small	☐	☐	☐	☐	☐	Large
Insensitive	☐	☐	☐	☐	☐	Sensitive
Colourless	☐	☐	☐	☐	☐	Colourful
Unsociable	☐	☐	☐	☐	☐	Sociable
Closed	☐	☐	☐	☐	☐	Open
Passive	☐	☐	☐	☐	☐	Active

For each of the following statements, please mark "X" in the box that best shows your response. Use the same 1-5 scale as shown earlier:

1: Totally disagree **2**: I disagree **3**: I neither agree nor disagree **4**: I agree **5**: I totally agree

1. I enjoyed doing the game tasks Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree

2. I felt nervous while using the map & phone system
Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree

3. I put a lot of effort into using the map & phone system
Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree

4. I think I am pretty good at using the map & phone system
Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree

5. I found using the map & phone system very interesting
Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree

6. I felt tense while doing the game tasks Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree

7. I think I did pretty well using the map & phone system tasks compared to others
Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree

8. I am satisfied with my performance at the game tasks
Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree

9. I tried very hard on the game tasks activities
Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree

10. I felt pretty skilled at the game tasks Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree

11. It was important to me to do well at the game tasks
Totally Disagree ☐ ☐ ☐ ☐ ☐ Totally Agree


Draw on the map which areas you think you walked over while in the game today

ORAL QUESTIONNAIRE FOR THE USERS

Researchers fill this: Researcher Name:

Team Name: Phone tag:

Questions:

How did you use the MapLens, can you show with a phone and a map? (Get them to demonstrate, ask if mind video-ing)

Did you know beforehand your team members? What relationship—friend, colleague, boss etc?

How was the experience?

(Choose which part of the question is appropriate to your interviewee)

Which parts did you take more time with? Which did you enjoy most? Which things related to game or technology, were you thinking more about /played more with / returned to or engaged most with.

Did pointing help you complete the map+phone tasks?

Did talking with the others help you complete the map+phone tasks?

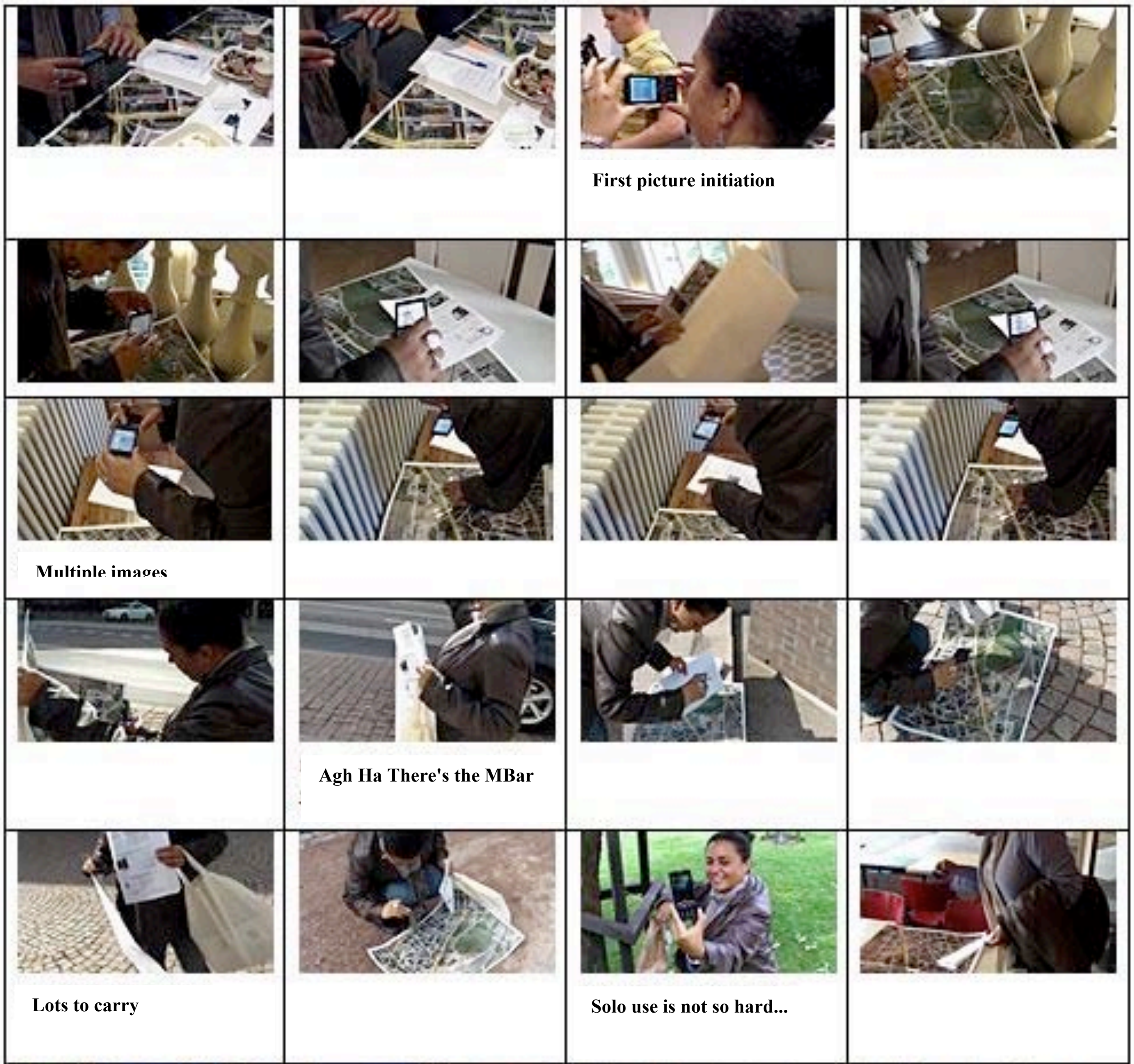
And then if the user said something interesting I would ask more about it, but letting the user speak as freely as possibly.

Appendix 2: Preliminary Video Observations of Trials

Solo users with solo device

Solo DEVICE: Solo USER

Trial Date: 23.8 Researcher:XXXX Team: XXXX



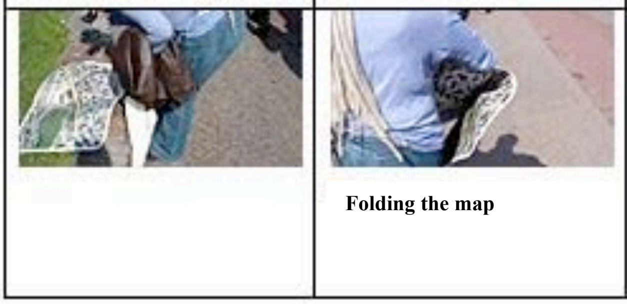
First picture initiation

Multiple images

Agh Ha There's the MBar

Lots to carry

Solo use is not so hard...



Folding the map

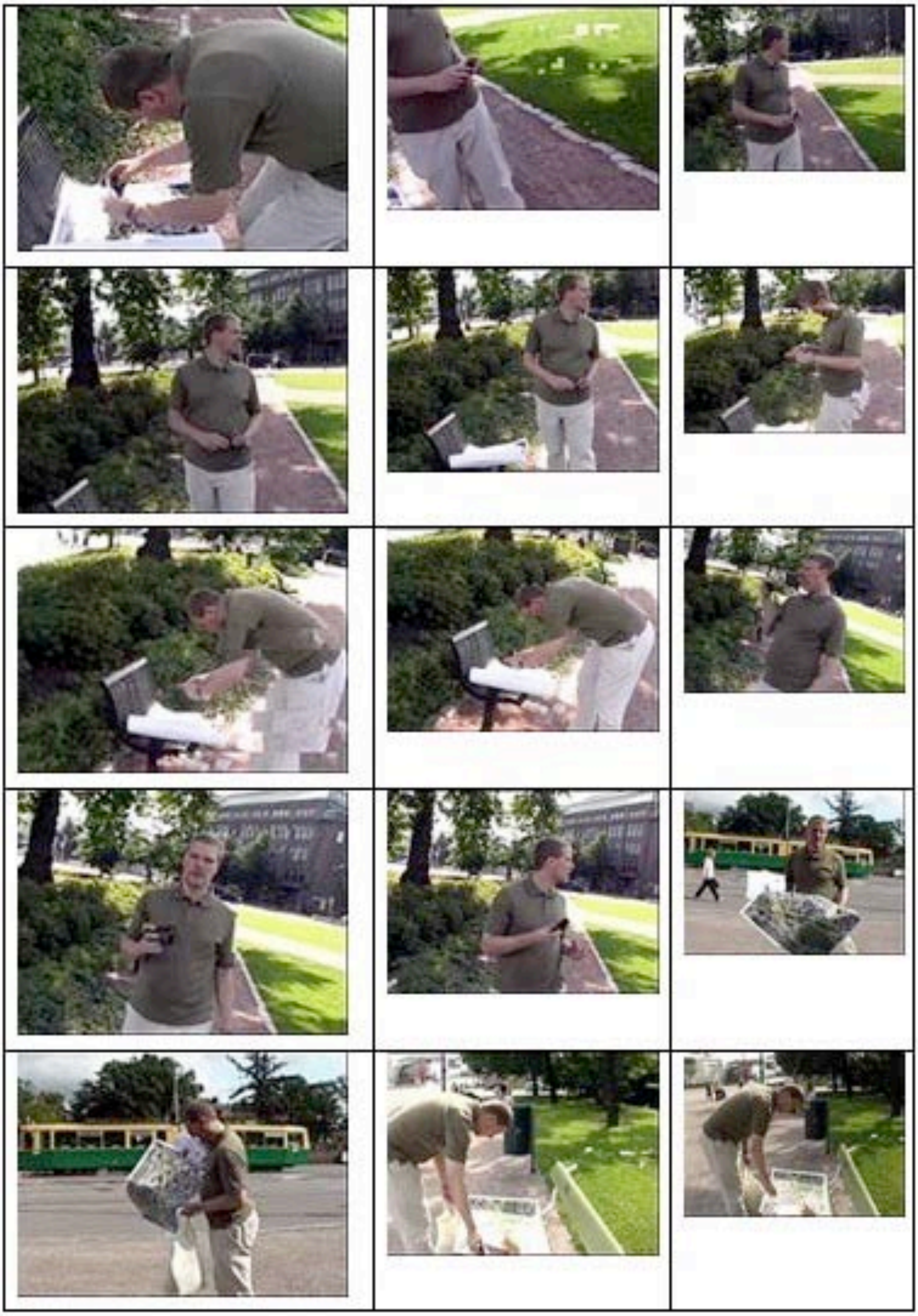
In the lobby she plans her route. Using a physical map when navigating, Kneeling down when uses ML usually. Not too organized with the kit, but used MP efficiently, found her way. Kept the application and lid open all the time. Using the “you are here” to locate herself, and worked methodically. One gesture to environment when sees MBar, and on pointing on map, and mid-air beneath ML. Browses other images and multiple images while in museum or inside Mbar, not while on move in outside.

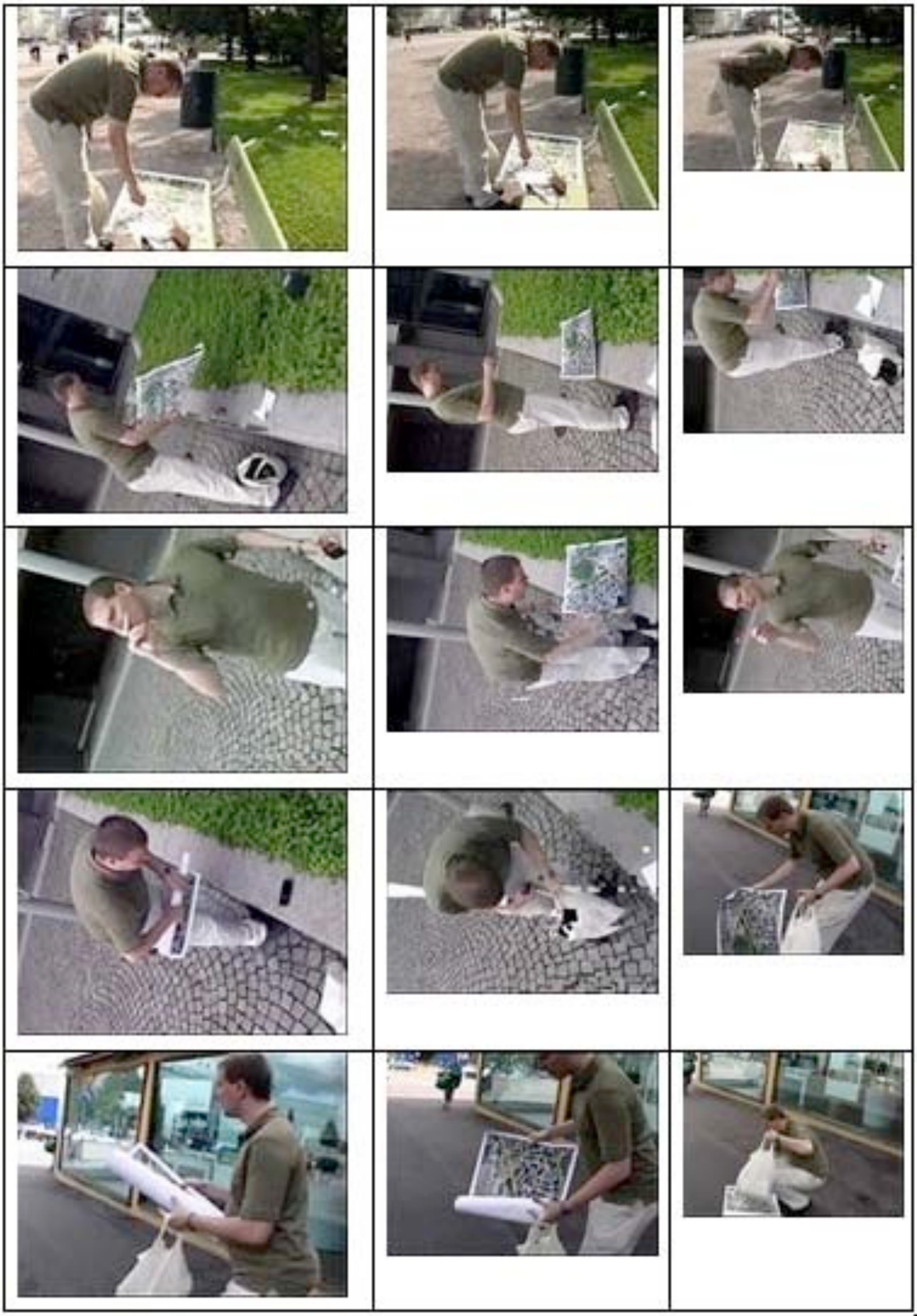
Solo DEVICE: Solo USER

Trial Date: 16.08 Researcher: XXX Team: XXXXX

Lots of online and offline use of others and own photos and clues (and enlarging images). One handed and two-handed use and largely horizontal, holding vertical while carrying. Skews device to body and map alignment. Stopping in street or walking using rolled up version of map. Looking from map and looking from device to the environment. Pointing on found area on map and looking up to the environment. Use of feet to control map. Managed with kit artefacts easily (and own backpack), keeping things close to body, clues book went inside bag after first using outside, good tidy system in place from the beginning. Strategy: he couldn't find fish in museum so came back later to find other users clues. Interacted with researcher and other teams.

SOLO USER: SOLO DEVICE
Trial Date: 23.08 Researcher: XXXX Team: XXXXX





One person team.

In museum just explored unorganized way. Outdoors navigated clue-by-clue, did not plan the whole route. MapLens was used very much the way demonstrated to participants in museum.
Carrying the device in pocket, device closed. He was not too organized with the kit. Once forgot clue booklet behind.

Adjusting his habits slowly when on go.

Did not use “you are here”. No pointing when using the system.

Pointing on map and screen only when explained something to the researcher or when trying to find something that he was having difficulty tracking. Also when used map to navigate without MapLens later in the event. Was going to use the device for the internet browser, but did not want to go out of the application. He is not too happy about the big map… Map rolled.

System of use. Usually puts map on bench or fence for using the system. Once kneeling on middle of the street and put map on ground. When found a good place to use a system used it there again. Mainly two-handed horizontal use, also one-handed vertical

Communicating with researcher: questions, facial expressions/gestures. In museum talking to other people.

“It would ne useful if you would be able to see later, what you have just seen by using MapLens” “It would be good to be able to see your own photos only.”

Teams sharing one device only

SHARE TEAM: One Device

Trial Date: 16.08 Researcher: XXXX Team: XXX : Phone G1 Share device

		Main AR user R, and put in pocket at times. However all used, shared screen, pointed under, pointed at screen, gestured to environment from device and map. He showed how to use as well. Folded map later in game. Pointing a lot on the map, with fingers, and R more gestures in the environment. Mainly one-handed use in environment. Some vertical use by brown haired gal, then pressed red button, and then she gets phone back and still vertical use. No comments or notice. He uses it vertically then horizontally, and two handed to one handed skewed and gestures to environment. Others knew way in city but they were guided by him with the tasks. Gave him some guided tour a bit as walking etc He was visiting couchsurfer and they were couchsurfer hosts and may know each other.

SHARE TEAM: One Device
Trial Date: 16.8 Researcher: XXX Team: XXXX

He was main user. And C 2ndary. He was carrying device all the time. When it did not work then it was switched to C and H. H learned to use it but did not really use it during the trial, did outside kmart. H was taking lot of photos. Kitbag & map were carried by all, mainly h and c. He got it back off Hwhen it stopped working. C holding a map, when he uses ML. Pointing on a map, lots of They seemed to be pointing only at the map, not on device because too distant the bodies, they didnt know each other and it impacted use, despite the tasks that attempt to disrupt this. Did not point at or use screen, as not close enough to see the screen of each other. so others kept out of it. Shy for physical proximity with him, not with each other.

He was not sharing the device, he was keeping it to himself, but he attempts to share with odd angle of screen. Did not interact with other teams. Interacted a bit with researcher for questions, photos. He asked people to take photos of them.

SHARE TEAM: One Device

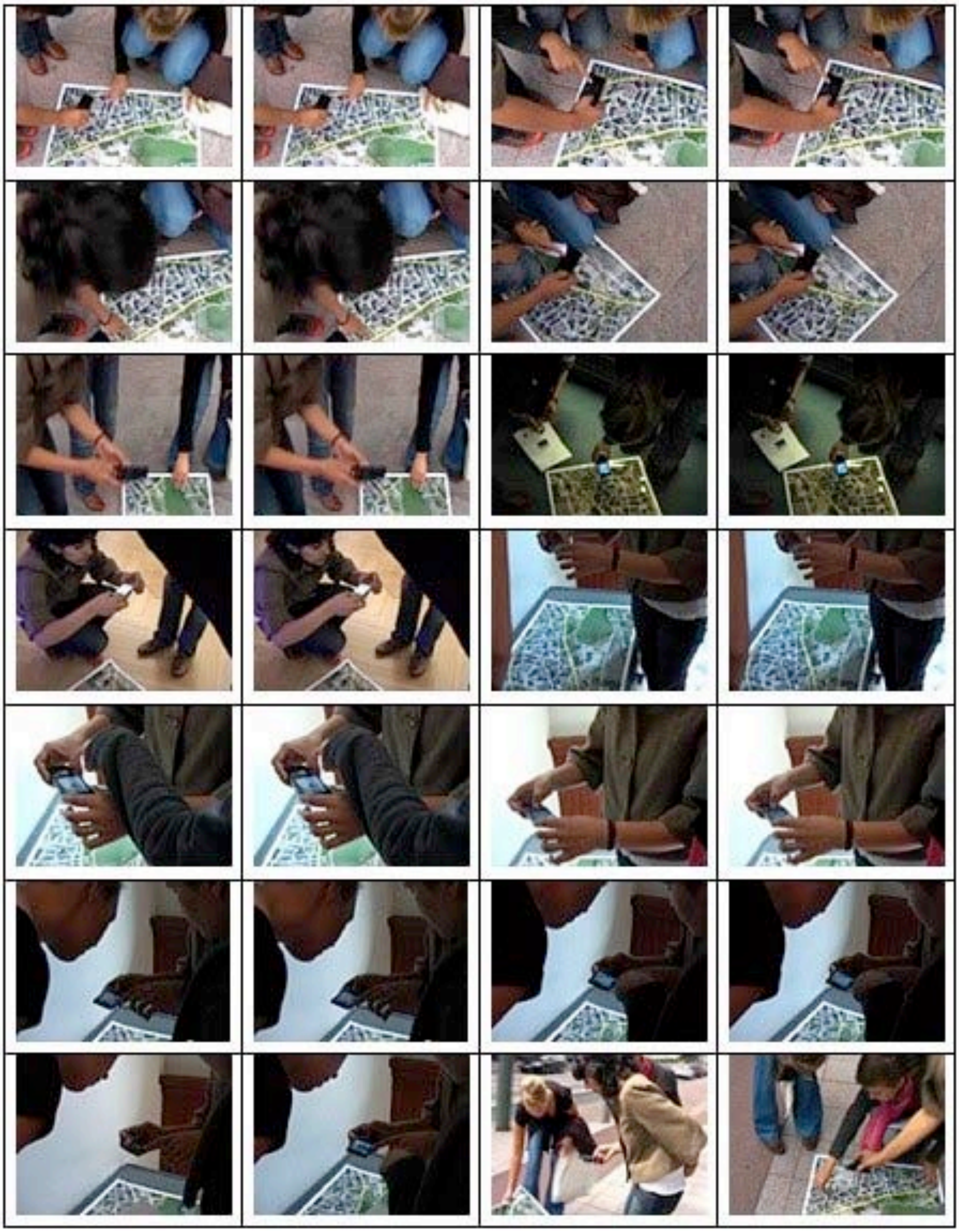
Trial Date: 16.8 Researcher: XXX Team Name: XXXX

<table>
<tr><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td colspan="2">BJ did most of map work at the beginning. She took the map most of the time.

Taking turns when using device BJ->blue guy -> girl. Pointing screen, map and environment and sharing device easily between two, sometimes three, pointing and heads close to see information. Much discussion with pointing to affirm on map before moving on. They found their photos online and looked those. 2 and 1 one handed use, mainly 1 handed. Phone usually horizontally. With other team they interacted.</td></tr>
</table>

SHARE TEAM: One Device

Trial Date: 23.08 Researcher: XXX Team: 2 XXXX



Offered turn taking. Closer to each other.
S carrying map, and kit. Switching roles easily. S who was outsider pointed other's screen bit less, stayed bit further etc.
MapLens main user was one L, and I as well. Easily switching phone between all team members. Usually all over the map.
Relaxed use when outside.
Map aligned with environment before starting. Map on ground or they were holding it. Pointing on map and screen. And clue

booklet. Also pointing environment. Using finger. Also device to environment and map. Pointing with a pen to screen & map. Gesturing on a map. They are looking the same screen, heads very close. Showing screen to each other. Also handing phone to each other. tilting screen to each other. One-handed horizontal, and one-handed vertically, also sometimes two-handed when difficulties, of when using buttons. Vertical for the holder, meant easier for others to see. aligned the phone to the others view. Nice sharing of the device. (Looks skewed).
Looking at gallery and navigating images it was two-handed use, otherwise often one-handed.

Working with map & device and then worked separately with just with map.
Took all the pictures themselves. Did not interact with researcher. Did not interact with spectators or other teams. Asked timer phone from researcher, did not use MapLens for that, time was seen as difficult from MapLens use.
Didn't want to leave MapLens. Preciousness that it might stop working, and particularly the first week, also researcher influenced.

Teams with three devices and three users

THREE DEVICES: THREE USERS
Trial Date: 16.08.2009 Researcher: XXXX Team: XXXX

Both using on floor in museum

Both using, he pointing under her lens and discussing

Two using on floor, other standing and using own phone browser to look for info on seal

Kate (right) switching between ML and own phone. Later Mitsu (left) uses own for GPS

Outside, Two on map, just quickly check on using available wall, don't put things down

One uses vertically and one pointing at screen, other at map, all help in use and discussing

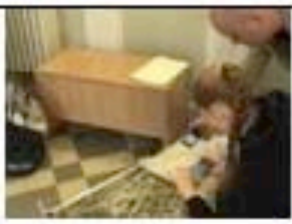
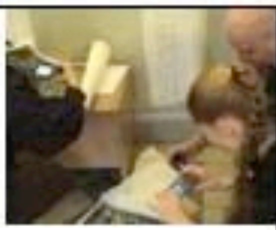
All work as team with pen, finger and one device

Standing up and solo use, while others water test, she finds next task location (sea water spot)

Marking on map the info for others

Use other devices and other uses of phone e.g. GPS, browser; vertical and horizontal use of device, largely two on map, share screen (not devices) pointing and circling with finger and pen and marking on map, matching finger under map, and pointing on screen and environment, divide tasks more individually. Fold map, working with boss. Switch attention from device to map, map to environment, device to environment. Largely one-handed use, two handed for selecting and navigating interface.

THREE DEVICES: THREE USERS
Trial Date: 16.08 Researcher: XXX Team: XXX

 Both trying on map on floor in museum.	 He changes position to find other information	 3rd member looks at own map	 Discussing between them all and their two maps and three phones
 All keep looking, 3rd at map below where info was augmented	 Share other map to discuss another aspect	 Outside with wind, first all try to use, then one folds map	 Then two hold map and one uses with all looking at the one screen
 She tries to use while walking	 She demos best way to use and aspects of short and long distance while walking (not with map)	 In MBar on table, both start using for next clue	 Then just one uses and skews phone around so both can see screen
 His no longer working from here (he is just still trying)	 On a short stop, Two confer and move on, put nothing down but stop walking for a pause in time.	 She reconnecting after battery change, use of two hands	 She using while walking to demonstrate an idea with phone for new environmental game task (without map)

All tried at first, showing screen to each other earlier in piece. First worked 2 maps, then outside one, folded as soon as hit the wind. Sharing device and pointing at screen and below on map. Tried walking, discussed and demo-ed best way to use while walking. Parked using aside from in e.g. museum, MBar where table. Short stops with 2 on device. Two handed use for e.g. connecting to elisa with first photo after battery change. Showing use for potential game on journey back. Device in hand from beginning to end. They switched the stuff between them easily. Working well as a team (e.g. all holding the map on first use). Experimenting how to use, and happy about it. Fluid switching of all stuff. Note: his stopped working, so then only two phones from Mbar. Many iconic gestures.

THREE DEVICES: THREE USERS

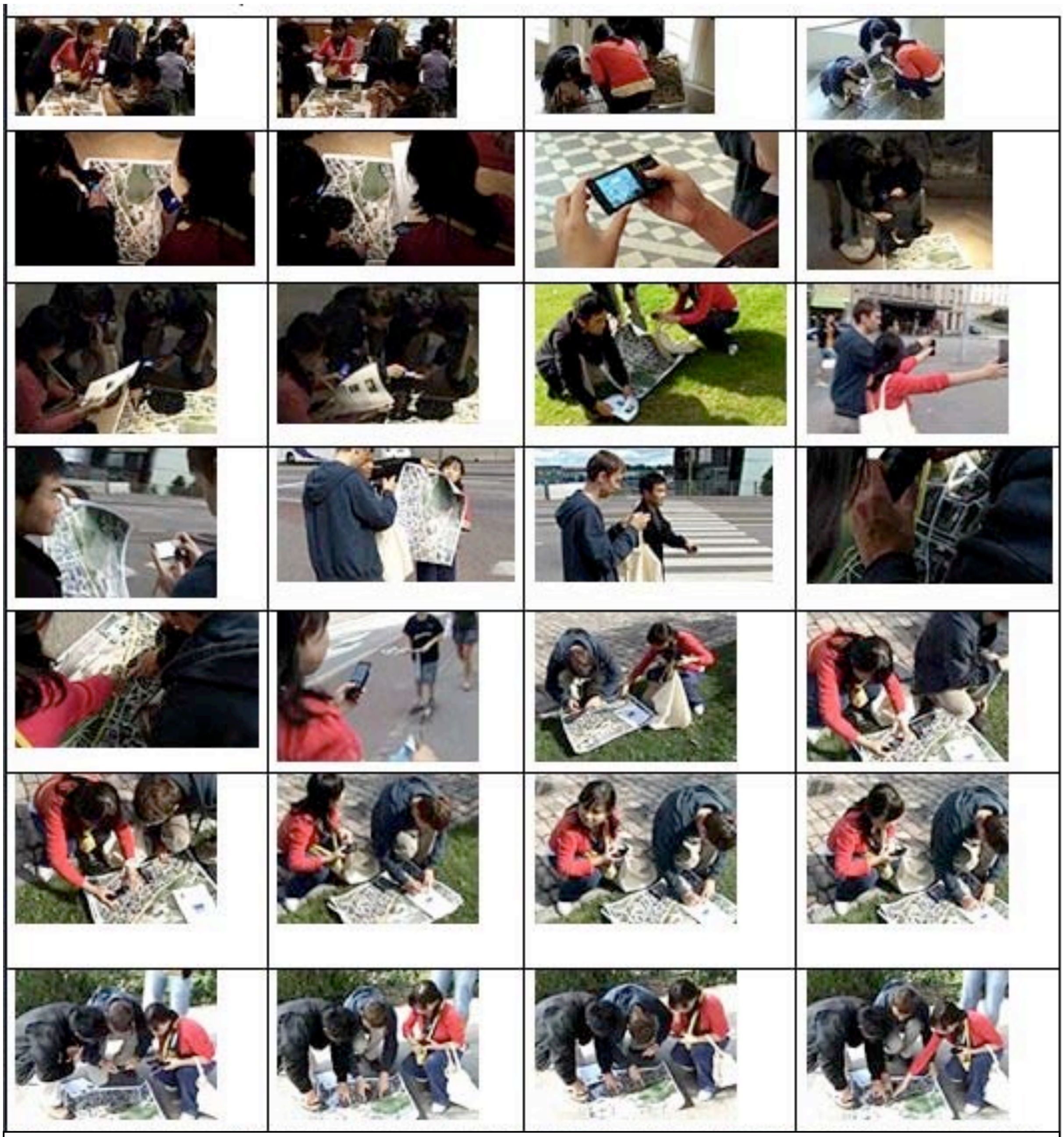
Trial Date: 16.8 Researcher: XXX Team: XXXX

Theme: She says: "It's here. We're going here, do you believe me?"

The girl (H) took hold of the map and orchestrated where to look from by pointing on the map (and to the environment). Guys augmented. Decisions made by S and H. S main photographer. H was the lead. Two handed, singlehanded and vertical & horizontal use. Shared device, H handed over hers when problems to leather jacket guy. Usually sharing one and S using his solo, cannot tell if his is working properly most of the time. She choreographed with pulling out map, pointing to the map, and the environment. Interacted with spectators, other teams and researcher. Starts with Laurel and Hardy type sequence of each trying to use own map outside in the wind. “1st_use.wmv”

THREE DEVICES: THREE USERS

Trial Date: 23.8 Researcher: XXX Team: XXX



G was taking care of booklet. B was hunting clues. B was photographer. M main augmenter. M = leader. They often had map on ground (also used vertical) and 1-3 of them using it. G used phone, but came back straight in so not sure if getting advice on where to go. All three often pointing on map before heading off. Sequence she looks from device to environment and swivels around (6th row). Gesturing in the environment. Hand mi-air (not on map) under device. G walks using browser. B: “Hurry up, c’mon, let’s go”, so they browsed multiples offline while walking. Pointing on each others screens, with map even if still looking swept off and on way. Lots of map pointing under device.

THREE DEVICES: THREE USERS

Trial Date: 23.8 Researchers: XXX Team: XXXX

	Two lead roles, girl passive as late call in, guy in black took first leading and expert role, then two boys co-lead or battled for leadership. Both used maplens concurrently on map. Roles defined from museum outwards, used on ground when started, then rolling map and dropping down, rolled map inwards system. Used ad hoc and on the move batteries, then it went awry because they separated and there on in stayed together. Grey guy used pen for clue book. All 3-person users using simultaneously on two occasions. Sharing, and looked through others, pointing on the others device, parallel use—“are you getting this?” Communicating while use in parallel. Two-handed for clicking through images, enlarging etc. and one handed for roving the map, standardly in one hand horizontal use Girl and Grey t-shirt Tried to use while walking.

THREE DEVICES: THREE USERS

Trial Date: 23.08 Researcher: XXXX Team: XXXX Phone C1, N4, N3

A kit (with other maps) and c map and phone and j only phone, C going solo, looking at map pointing, looking at environment, walking, others lost dragging map open, J clue booklet and A 2 maps in bag, search for surface and A uses phones, C leaves map, they leave map there a its her map, do battery task, and he searchers for batteries throw out maps on ground, interacting with researcher for photo, J's role at some points, he went to gallery to see photos they had taken. J then responsible for pictures and wins prize. J rolls up all maps. Mbar vouchers , so A responsible for kit and C now taken over clue booklet.
Walking with phones vertical. J takes photo of man standing on a bench, others not that interested. Took photo of G (solo user) but did not interact directly.
J took over sunleaf photo. And clue booklet for last sunlight task in toilet. C looking at watch. Finishing task. Pointing with finger and pen on map. One-handed use horizontal, two-handed C with clicking through and pen and clue book in hand. J used MapLens while walking against the environment.

Acknowledgements and Further Information

IPCity is partially funded by the European Commission as part of the sixth framework (FP6-2004-IST-4-27571

For further information regarding the IPCity project please visit the project web site at:

ipcity.eu