D4.3.2

RESULTS AND FEEDBACK ANALYSIS - FINAL

End user feedbacks after 2 iterations

April 2015

ABSTRACT

This deliverable presents the results and end user feedbacks delivered by Zurich, and Barcelona Experimentation Sites after two evaluation cycles in the context of the Pervasive Game Platform of the FI-CONTENT 2 project.

This document is a deliverable of the FI-CONTENT 2 integrated project supported by the European Commission under its FP7 research funding programme, and contributes to the FI-PPP (Future Internet Public Private Partnership) initiative.

DISCLAIMER

DELIVERABLE DETAILS

[Full project title]: Future media Internet for large-scale CONTent experimENTation 2
[Short project title]: FI-CONTENT 2
[Contract number]: 603662

[WP n°]: WP4: Pervasive Games platform
[WP leader]: Stéphane Magnenat, DRZ

[Deliverable n°]: D4.3.2
[Deliverable title]: Results and Feedback analysis - Final
[Deliverable nature]: Report (R)
[Dissemination level]: Public (PU)
[Contractual delivery date]: M25 -April 2015
[Actual delivery date]: 12 May 2015
[Editor]: Nathan Gouveia (GOBO)
[Internal Reviewers]: Stéphane Magnenat, DRZ, Marcel Lancelle, ETHZ
[Keywords]: Experimentation Sites, Evaluation
[File name]: FI-CONTENT2_WP4-009_D4.3.2_V1.0.docx

EXECUTIVE SUMMARY

This document describes in summary form the successes of the experimentation of technologies developed in the context of the Pervasive Game Platform of the FI-CONTENT 2 project, suggestions of improvements and conclusions.

LIST OF AUTHORS

Organisation	Author
DRZ	Stéphane Magnenat, Mattia Ryffel
i2CAT	Marc Aguilar, Pau Pamplona, Sergi Fernández
ETHZ	Marcel Lancelle, Fabio Zünd
DNET	Boris Pokric
BLRK (Disney UK)	Kenny Mitchell
Studio Gobo Ltd	Nathan Gouveia

TABLE OF CONTENTS

ABBREVIATIONS

AR	Augmented Reality
CG	Computer Graphics
FI	Future Internet
FI-PPP	Future Internet – Public Private Partnership
GPS	Global Positioning System
GPU	Graphics Processing Unit
HCI	Human Computer Interaction
iOS	Apple iDevice Operating System
LED	Light Emitting Diode
POI	Point of Interest
SE	Specific Enabler
SLAM	Simultaneous Localization and Mapping
GE	Generic Enabler

1 - INTRODUCTION: SUMMARY OF THE EXPERIMENTATION CYCLES

1.1 - Purpose of this document

The objective of this document is to provide a report on experimentation results and user feedback on the technology released as part of the Pervasive Game Platform, as part of Work Package 4 of the FI-CONTENT 2 project. The document is final, in that it gathers data and feedback from the first and second experimentation cycles.

1.2 - Reading this document

This document summarises each of the WP4 technologies and experiments. It describes the experimentation, what worked well, how improvements could be made in the future, whether it met its objectives, and provides a conclusion.

This document is labelled D4.3.2, and is part of a number of similar documents released by the project. Three groups of deliverables are relevant when reading this document:

The first group includes D2.3.1, D3.3.1 and D4.3.1. These documents are the summary of the first experimentation cycle with respect to the three content platforms of the FI-CONTENT project. A common structure was created to present content in a unified way.

The second group of relevant deliverables includes D7.1.2, D7.5.1 and D7.6.1, which collect the outcome of the user experimentation sites in the first experimentation cycle, and in particular the ones in Barcelona and Zurich. In the first experimentation cycle, all experiments in relation to the Pervasive Game Platform have been conducted on these two sites. To avoid duplicating the content, this document will include only a summary of the experiments performed plus information that is general to the Pervasive Gaming Platform. The reader is invited to inspect these other deliverables where more details are presented.

The third group of relevant deliverables include D7.5.2, D7.6.2 and D4.1.3, which collect the outcome of the user experimentation sites in the second experimentation cycle, and in particular the ones in Barcelona and Zurich. In the second experimentation cycle, all experiments in relation to the Pervasive Game Platform have been conducted on these two sites. To avoid duplicating the content, this document will include only a summary of the experiments performed plus information that is general to the Pervasive Gaming Platform. The reader is invited to inspect these other deliverables where more details are presented.

2 - FIRST EXPERIMENTATION CYCLE

2.1 - Overview of the scenarios for experimentation

In the context of the Pervasive Games Platform, we have conducted a number of experiments involving users, asking them to test and evaluate applications and technology components, in order to gather feedback to steer further development and extension of the platform.

Table 1 includes the list of experiments in chronological order as they were run in the Zurich and Barcelona Experimentation Sites.

Table 1 Pervasive Game Platform Scenarios tested in the first experimentation cycle

Date	Site	Scenario	Tier
April 2013	Barcelona	1. Attraction Driving Content Sharing	2
July 2013		2. Augmented Reality in the Wild	1
November 2013		3. Seamless Augmented Reality in the Web	1
November 2013	Barcelona	4. Virtual Character Synchronization on the Web	1
December 2013	Barcelona	5. Tabletop Augmented Reality Games	1
February 2014	Barcelona	6. Immersive Control Systems	2

As mentioned in D4.1, the Pervasive Games Platform targets three tiers of complexity, namely:

1. Augmented Reality Toys
2. Installation Based Applications
3. City-Wide Games

As indicated in the table, this first experimentation cycle touched the first two, while the city-wide games experiments were conducted in the second experimentation cycle. This choice was purely pragmatic, as the three tiers are increasingly more complex, both in terms of the technology required and the logistic for deploying and testing them.

2.2 - User Centric Approach

Our experiments follow a user centric design, which has been applied as guiding principle during all phases of the creation of Pervasive Games Platform. As a brief summary, we have considered the end users' needs and interests in domains where FI-PPP technology would meet our use case, i.e. media content related to gaming, augmented reality, and pervasive experience, and have designed a number of scenarios, described in D4.1, which we believe are interesting for the public and can generate new business avenues. These scenarios have then driven the creation of a number of technology components, also called enablers, released in September 2013 as part of the Pervasive Games Platform, described in D4.2. Finally, the components have been combined to create applications that reflect the original scenarios, and they have been presented to the users.

In practice, these three steps were intertwined, as we have progressively designed scenarios, built technology and started experiments with users from early on in the project until now. What matters though is that the question of whether or not we are going in the right direction is answered at the experimentation step, which acts as check point for the quality of the work done so far. So, in the first experimentation cycle, we have gathered user feedback, with the main purpose of validating the scenarios as well as the technical components and applications we have built. The main line of questioning is whether or not the scenarios are interesting for the users, and whether the technology provided actually support the scenarios in a satisfactory manner.

2.3 - Cooperation between different sites

Cooperation between different sites has three main purposes. First, each site has unique access to local communities, infrastructures, and cultural heritage. Different culture may react differently to technology and applications; hence, feedback may present interesting differences. Second, validation across multiple sites strengthens the value of the result. Third, experiments may actually involve interaction of services and users at different locations.

In this first experimentation cycle, the main collaboration happened between Barcelona and Zurich experimentation sites. For a detailed description of these two sites, refer to D7.1.2, Section 5 and 6 respectively. As mentioned before, the experiments touched Tier 1 and 2, and were executed individually on the different sites.

All experiments were conceptualized and developed as team effort with contributions from all partners. For practical reasons, the actual deployment and testing was done in Zurich first. The presence of both DRZ and ETHZ technical staff facilitated the execution of the earlier prototypes, which required some expertise to be run. Prototypes have been refined afterwards to allow portability to dissemination events, allowing testing in other sites. In the second round of experiments, the quantitative analysis of the Gnome Trader application took place in Barcelona with i2CAT in charge of the operations.

3 - Second experimentation cycle

3.1 - Overview on Scenarios

The scenarios tested in the second experimentation cycle in Zurich and Barcelona included prototype games from tier 3 (city-wide game) scenarios. In addition, we refined the evaluation of some tier 1 (toys) and tier 2 (location based) scenarios.

Table 2 Pervasive Game Platform Scenarios tested in the second experimentation cycle

Date	Site	Scenario	Tier
June 2013 to Oct. 2014	Zurich	Tabletop AR games	1
December 2014	Zurich	Creating SLAM-based AR Game Using Web Creator Tool	2
April 2013 to July 2014	Zurich	Attractions Driving Content Sharing	2
July 2014	Zurich	Location based Virtual Reality	2
July 2013 to Sept. 2013	Zurich	Augmented Reality in the Wild	3
Sept. 2014 (Zurich), Dec. 2014 (Barcelona)	Zurich, Barcelona	City-Wide Economic Game	3
Sept. 2014 to Dec. 2014	Zurich	City-Wide Scavenger Hunt Game	3
December 2014	Zurich	City-Wide Augmented Strategy Game	3

3.2 - User Centric approach and qualitative and quantitative methods

As planned, and stressed by the reviewers at the first review meeting, we conducted thorough user studies following strict scientific protocols, which lead to several publications in international peer-reviewed venues. This is an external validation of the soundness of our evaluation methods.

We combined qualitative and quantitative methods, aiming at measuring both the efficiency of the SEs and their added value for the users.

3.3 - Cooperation between different sites

The scenario “City-Wide Economic Game” is evaluated jointly by the Zurich and the Barcelona experimentation sites.

4 - EXPERIMENTATIONS

In this Section, we present details about the experimentation of the scenarios listed in Table 1 and Table 2. Notice that, more detailed information is presented in deliverables D7.5.1, D7.6.1, D7.5.2 and D7.6.2. Here, we summarize the experiments and put them in context of the platform.

4.1 - Scenario “Attraction Driving Content Sharing” by running “ARPix”

4.1.1 - Description of tested application

In this scenario, we explore a case of installation-based application (tier 2) aiming at increasing the entertainment value of a venue.

The ARPix application uses a billboard as marker and allows the users to take augmented reality pictures together with a virtual character. For a movie theatre, characters from the movie posters or cut-out cardboard can be added into the user picture.

4.1.2 - Test objectives and expected outcomes

The main goal of this experiment is to evaluate the robustness of the Reflection Mapping SE and Camera Artefact Rendering SE and to validate whether they improve the user experience. We were also very interested in gathering feedback from the users about the application concept, and see their reactions to the pictures created.

4.1.3 - Applied methods and tools for evaluation

In the first experimentation cycle, we briefly interviewed parents and their children attending a Zurich movie theatre. We also analysed transient video footage of the guests enjoying the apps and their interactions with the handheld AR devices.

Figure 1

In the second experimentation cycle, we presented users with four pictures with the enablers on and off, and stored the preferred image index for each picture. This choice was seemly integrated into the user interface of the tool, making it a fully user-centric method. In addition, we stored pictures made by the users.

4.1.4 - Summary of Experimentation in Zurich

In the first experimentation cycle, a first internal experiment took place at Disney Science-Fair, March 2013. It was followed by an official public playtest at the Arena Cinema in Sihl-City Zurich on April 5th 2013. A particularly crafted poster was placed in an area with enough space for people to stand and take pictures. Local visitors of the theatre were invited to participate to the experiment as volunteers. Participants would pose in front of the poster and a family member or a team member of DRZ or BLRK would take the camera shot.

In the second experimentation cycle, we conducted a quantitative experiment at two different locations (1: 32 users and 2: 40 users), with the aim of validating the contribution of the enablers to the user experience. Participants were of different ages, genders and educational backgrounds, and similar lighting conditions.

The participants at location 1 were less experienced in visual computing and games as the participants at location 2. The experiment took place on July 9-10, 2014. In addition, ARPix was deployed in the Grand Central Terminal in New York City.

The Reflection Mapping SE has been created as collaboration between BLRK and DFKI. The Reality Mixer - Camera Artifacts Rendering SE has been created by ETHZ. The tested application was developed by BLRK and DRZ, and both partners participated in the field testing in Zurich. The quantitative experiment was run by ETHZ and DRZ, while the NYC deployment was conducted by DRZ.

In the quantitative experiment during the second experimentation cycle, in both locations the participants mainly voted for the image with correct lighting and enabled camera artefact rendering. A Chi-square test confirms a significant difference in groups for both locations. In location 2 a greater part of the participants voted for configuration B compared to location 1. We believe that this is due to the fact that the percentage of technical people that are working in the field of visual computing was significantly higher in location 2 than in location 1. Participants with a visual computing background may be more sensitive to image effects and may recognize the slightly blurry virtual content and consider it a visual defect, whereas less technical participants would not directly spot the blur but unconsciously feel that the image with camera artefact rendering enabled blends better into the background. At the Grand Central Terminal at NYC, 700 pictures were taken, with an average of 2 to 3 people per picture, sometimes more. This lead to 2000 users interacting with ARPix.

4.1.5 - How improvements could be made in the future?

The pilot experiments led to further ideas for enhancements to the experience, such as, better interactive photo button placement, bigger buttons for children, a mirror view to take self-portraits, higher resolution and more seamless integration of the photo with use of the Reality Mixer - Camera Artefact Rendering enabler concept. Some of these ideas were implemented for the tests in the second experimentation cycle. In the future, more visual aspects of the reality could be captured, and more detailed models of camera artefacts rendering could be implemented, leading to even more realist shading of CG content and therefore better integration of the augmented content in the reality.

4.1.6 - Did it meet it objectives?

The pilot experiments in the first cycle showed that the overall assessed engagement and enjoyment was very high and guests asked when the app would be publicly available. The quantitative experiments in the second cycle confirmed that both enablers improve the user experience significantly. In addition, the large scale NYC deployment proved the strong interest from the general public for this immersive experience.

4.1.7 - Summary of outcomes and Conclusion

Through the Reflection Mapping SE and Camera Artefact Rendering SE enablers, we investigated the impact of visual realism in mobile AR for the category of applications where aesthetic quality is a factor. Based on quantitative analysis, we found a strong preference to the inclusion of more realistic lighting environment matching, an approach that our two enablers make it accessible to third-parties developers.

4.2 - Scenario "Augmented Reality in the Wild" by running "Skye Wars"

4.2.1 - Description of tested application

This experiment is labelled Tier 1, as in the end it is a simple application running on a device without connectivity. However, it is our first attempt to enable AR tracking in an unstructured environment, which is a necessary step to move toward Tier 3 and city-wide gaming.

The Skye Wars application was developed targeting the computer graphics conference *SIGGRAPH 2013*. During the conference the robot balloon Skye Blimp [5] was flying around in the hall before movie screenings. Participants could download the application and by following the blimp with the phone they were able to see an epic space battle happening around it.

Figure 2

4.2.2 - Test objectives and expected outcomes

The main objectives were to gather feedback and reactions from the users, to evaluate the robustness of the Fast Feature Tracking SE in a non-controlled environment, and in turn validate the concept of combining Fast Feature Tracking SE and gravity based AR (from the device gyros) for Tier 3 applications.

4.2.3 - Applied methods and tools for evaluation

Participants were informally interviewed during and after the event.

4.2.4 - Summary of Experimentation in Zurich

4.2.4.1 - Dates (Start/end) of the experiment

During the SIGGRAPH graphics conference that took place in Anaheim from 21th to 25th July 2013.

4.2.4.2 - Role (involvement) of partners

The Fast Feature Tracking SE has been created by BLRK. The tested application was developed by BLRK and DRZ, and both partners participated in the field testing, demo setup and support.

4.2.4.3 - Short Report

During the computer graphics conference participants were provided with a link to the apple app store to download the application and some simple instructions. Participants had to point their phones to the blimp balloon and hit the start button of the app. Then the players could shoot the enemy AR star ships with the Skye laser and see who can get the highest score.

Most of the targeted end users were researchers, software developers or artists; the number of users was around 1500.

During and after the event participants were informally interviewed by DRZ and BLKR experimenters. As the crowd at the presentation was quite big we had a lot of interesting feedback from the participants.

4.2.5 - What worked well?

The Augmented Reality Fast Feature Tracking enabler performed surprisingly robustly, given the lighting conditions in the Arena. Many users reported an enjoyable experience and recall this as a good year for the traditional interactive audience experience at SIGGRAPH. The game experience was entirely unhindered by the processing of the enabler allowing rich content to be simulated.

4.2.6 - How improvements could be made in the future?

The enabler would apply equally well to AR glasses or other display methods and this would be expected as these related technologies develop. As a color feature tracking method, the range of features could be extended or applied to compose tracking of multiple connected color features.

4.2.7 - Did it meet it objectives?

The objective of realizing Tier 3 application of tracking in the wild was perhaps most challenging scenario to execute well. The associated enabler exceeded the objective in terms of robustness and applicability also to Tier 1 and Tier 2 scenarios. The Skye Wars app (first app of FI-PPP to be released on the AppStore) is rated 4+ accessible to everyone.

4.2.8 - Summary of outcomes and Conclusion

Feedback was very positive and complimentary given the technical knowledge of guests and there awareness of harsh lighting conditions for robust operation of the Fast Feature Tracking enabler, although it was not clear that participants understood the app could actually be played anywhere away from the robotic Skye Blimp, which provided the focus of the event, but was not critical to enjoyment of the game.

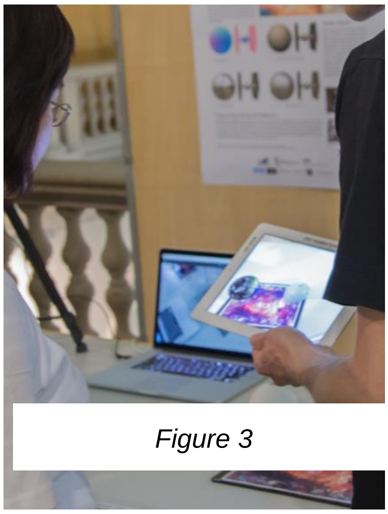


Figure 3

All the participants interviewed were able to run the application and track the blimp successfully.

This experiment was indeed a success. We were able to verify the robust functionality of the Fast Feature Tracking SE.

4.3 - Scenario "Seamless Augmented Reality in the Web"

4.3.1 - Description of tested application

The *Star Tours* application showcases the Reality Mixer - Reflection Mapping SE. Spheres placed at a certain position on top of the image marker are used to include as lighting probes to integrate realistic lighting effects for virtual objects in a very efficient way.

4.3.2 - Test objectives and expected outcomes

While testing the robustness and operation of the reflection mapping technique, we were also interested in the feedback from the users. Especially we were interested to see how much this technology improves the immersion and realism of the virtually rendered objects and if it justifies the additional necessary sphere(s).

4.3.3 - Applied methods and tools for evaluation

Informal feedback was acquired during and after the experimentation.

4.3.4 - Summary of Experimentation in Zurich

4.3.4.1 - Dates (Start/end) of the experiment

The Star tours demo was presented and tested by interested visitors at the Swiss Vision Day 2013, NEM 2013 and CeBIT 2014.

4.3.4.2 - Role (involvement) of partners

The Reflection Mapping SE has been created as collaboration between BLRK and DFKI. The Camera Artefacts Rendering SE has been created by ETHZ. The tested application was developed by BLRK and DRZ, and both partners participated in the field testing in Zurich.

4.3.4.3 - Short Report

The application was demonstrated to potential users who could then try on their own, moving the iPad to view the virtual objects from a different location. The experimenters also used flash lights to better demonstrate the dynamic lighting effects on the virtual objects, captured by the lighting probes through the camera of the tablet.

4.3.5 - Summary of outcomes and Conclusion

Many people found the technology very interesting and some were surprised about the low computational cost. A few people noted that the virtual objects were too small and thus, the effect was not so easy to see.

The operation of the Reality Mixer - Reflection Mapping SE proved successful and led to integration into the Augmented Resistance game.

4.4 - Scenario "Virtual Character Synchronization on the Web" by running "Spider Game Demo"

4.4.1 - Description of tested application

In the Spider Game Demo, a picture on a table or ground is augmented with a virtual simulated spider character crawling on it. Multiple players, each having a unique view of the game through their handheld devices, share the state of the spider.

The application takes advantage of today's web technologies on mobile devices to augment the camera image based on the marker, mix it with a virtual character and perform efficient scene updates across multiple devices.

Figure 4

4.4.2 - Test objectives and expected outcomes

The purpose of the envisioned experimentation was twofold:

- Generate professional user feedback on the application and its utilised enablers
- Raise awareness among stakeholders and specifically within the developer community of the FI-CONTENT2 project's offering of technological enablers and their possibilities

4.4.3 - Applied methods and tools for evaluation

The evaluation plan for the experiments with this scenario involved two questionnaires. One questionnaire was prepared to elicit the participants' feedback on the application, from the point of view of the end user playing with the game. The other questionnaire was geared towards getting the developers' opinion on the underlying enablers that made the application possible, from the perspective of the professional evaluating the functioning and possibilities of the game prototype.
In the Zurich experimentations a more informal method was applied. We cared especially about the feedback and the impressions of the participant's right after the playtest. We observed the participants reactions during the playtest and we carried out brief informal interviews during and after the playtests.

4.4.4 - Summary of Experimentation in Zurich

The full details of the experimentation with this scenario in the Zurich site can be found in D7.6.1.

4.4.4.1 - <u>Dates (Start/end) of the experiment</u>

The Spider demo was presented and tested by interested visitors at the Swiss Vision Day 2013, NEM 2013 and CeBIT 2014.

4.4.4.2 - <u>Role (involvement) of partners</u>

The tested application was developed by DRZ and DFKI. DRZ developed the iOS application while DFKI provided the xml3d website. DRZ, DFKI, ETH and BLRK participated in the field testing in Zurich and helped with the demo setup and support.

4.4.4.3 - <u>Short Report</u>

A booth was prepared to host the demo. It consists in a screen showing the xml3d webpage with an aerial view of the scene and few tablets that provide the AR view of the game and are able to shoot rockets at the spider.

After a short presentation of the app and the technologies behind it participants could experiment on their own with the app. During and after the playtest they were interviewed and asked for feedback. Participants overall seemed to enjoy the game. The application ease of use is really straightforward; the team had to provide very little support. Participants could just pick up the tablet and start shooting the spider while others could watch the game from the xml3d webpage.

4.4.5 - Summary of Experimentation in Barcelona

The full details of the experimentation with this scenario in the Barcelona site can be found in D7.5.1.

4.4.5.1 - Dates (Start/end) of the experiment

The Pervasive gaming platform experiments in this first experimentation cycle were conceived as a set of combined demo workshops of the Spider Game demo and the Augmented Resistance game, running over three phases, from the 6th through the 25th of February. For this reason, most of the descriptions contained in the section below also apply to the experimentation with the scenario “Tabletop Augmented Reality Games” in chapter 4 of this report.

The specific actions performed with the dates and numbers of users involved are listed in Table 3 below:

Table 3 Summary of actions in Pervasive gaming demo workshops

Name of demo workshop	Date	Users involved
I AM workshop	6th February	15-20 AR workshop attendees
BCN Lab	21st February	15-20 event attendees
AR-xperiment (2 sessions)	24th & 25th February	10 + 10 game developers

4.4.5.2 - Role (involvement) of partners

The technical setup and preparation of the demos was carried out in close coordination with the WP4 partners, more specifically DFKI, DRZ and ETH Zurich. Also, Mr Chino Noris of DRZ conducted a presentation and brief training session during both sessions of the AR-xperiment, to be further explained in section 3.4.5.3.

For the organisation of the event in the Fabra & Coats building, I2CAT also collaborated directly with Sigma Orionis.

4.4.5.3 - Short Report

The purpose of the initial demo at the I AM workshop was to fine-tune the technical setup of the demos and improve the feedback questionnaires. The initial feedback received from users provided a first assessment of the gaming demos, and allowed the refinement of the questionnaires with a set of more specific and deeper-probing questions.

This was followed by a larger, real-world demo workshop at the BCNLab event. The BCNLab demo was a response to an emerging opportunity to disseminate the project’s enablers and gather feedback from a professional audience, composed of multidisciplinary artists, creative industries representatives, and ICT officials from city councils across Europe.

Finally, the AR-xperiment was a two-day training and testing session with two groups of 10 professional developers. These sessions lasted approximately two hours and a half each (see Table 8 below), and consisted in an introductory presentation and brief training on the enablers delivered via videoconference (Bluejeans) by Mr Chino Noris, and a hands-on session in which the attending developers were encouraged to try and test both applications themselves. At the end of the session, attendees were asked to fill in two feedback forms for each tested application, providing their detailed opinion on these prototypes.

4.4.6 - **Summary of outcomes and Conclusion**

The main recommendations that arise from the experimentation in this scenario in the Barcelona site are:

1. More functionality should be added to turn this very promising demo into an attractive and addictive AR game.
2. Specifically, the development of more game modes, options and actions offers possibilities to create attractive gameplay elements which make Augmented Reality an integral element of the game experience.
3. Exciting gameplay dynamics that users considered interesting for this game are enabling coopetition strategies (users cooperate to shoot at high-value targets, but end up with more or less points depending on their ability to defect at the right moment), enabling combo hits, having the spider shoot back at users (so that they must move to avoid damage), and adding more enemies with different score values to give rise to different game-winning strategies (focusing on few high-value, high-profile values or many easier low-value targets).
4. The visual elements of the terrain could also be integrated into the game experience, with permanent effects of explosions on landscape, and the spider being able to hide behind rocks, trees and other elements.

Finally, from the developer point of view, the existing documentation should be enhanced with step-by-step tutorial for beginners, more images and diagrams, source code examples, video demos, and a downloadable mock project.

4.5 - Scenario "Tabletop Augmented Reality Games" by running "Augmented Resistance"

4.5.1 - Description of tested application

In the Augmented Resistance demo, a physical tower on the center of a board must be defended against hordes of virtual characters in augmented reality.

Figure 5

4.5.2 - Test objectives and expected outcomes

In this application, the augmentation of traditional board games with the help of mobile devices is explored. AR Tracking is used to situate the device with respect to the board. Real objects are mixed with virtual ones. A light-probe system is used to capture the light of the environment, and uses it to illuminate the virtual content, obtaining a better matching of the virtual elements to the real ones.

The purpose of the envisioned experimentation was twofold:

- Generate professional user feedback on the application and its utilised enablers
- Raise awareness among stakeholders and specifically within the developer community of the FI-CONTENT2 project's offering of technological enablers and their possibilities.

4.5.3 - Applied methods and tools for evaluation

The evaluation plan for the experiments with this scenario involved two questionnaires. One questionnaire was prepared to elicit the participants' feedback on the application, from the point of view of the end user playing with the game. The other questionnaire was geared towards getting the developers' opinion on the underlying enablers that made the application possible, from the perspective of the professional evaluating the functioning and possibilities of the game prototype.

In the Zurich experimentations a more informal method was applied. We cared especially about the feedback and the impressions of the participants' right after the playtest. We observed the participants reactions during the playtest and we carried out brief informal interviews during and after the playtests.

4.5.4 - Summary of Experimentation in Zurich

The Spider demo, a preliminary version of this scenario, was presented and tested by interested visitors at the Swiss Vision Day 2013, NEM 2013 and CeBIT 2014. Also few informal playtests were done internally by DRZ where people visiting our lab could try the game. In both these events a booth was set up to host the demos. The setup consists in a few tablets and a game board with a 3d printed base stand where a ball is placed. First a brief presentation of the app is given to the public, and then the participants are given the opportunity to play with the app. Participants were observed in general during the play test and a few people were briefly interviewed for informal feedback.

Augmented Resistance app was developed by DRZ. DRZ, ETH and BLRK contributed in the development of the SEs used by the application. All partners helped the setup of the demo and participated in the field testing.

4.5.5 - Summary of Experimentation in Barcelona

The Pervasive gaming platform experiments in this first experimentation cycle were conceived as a set of combined demo workshops of the Spider Game demo and the Augmented Resistance game, running over three phases, from the 6th through the 25th of February.

The technical setup and preparation of the demos was carried out in close coordination with the WP4 partners, more specifically DFKI, DRZ and ETH Zurich. Also, Mr Chino Noris of DRZ conducted a presentation and brief training session during both sessions of the AR-xperiment. For the organisation of the event in the Fabra & Coats building, I2CAT also collaborated directly with Sigma Orionis.

This experiment was carried out simultaneously with the aforementioned experimentation of scenario "Virtual Character Synchronisation on the Web". A detailed account of the running of the experimentation can be found in section 3.3.1 of deliverable D7.5.1, and in section 4.3 -of the present report.

4.5.6 - What worked well?

Most of the feedback came spontaneously from the participants who seemed to enjoy the game and started competing for the highest score thanks to the Leaderboard SE implementation from ETH. Also worth saying that some participants started playing with the Reflection Mapping SE by pointing LED lamplights or their cell phones lights to the light probe ball. In general people enjoyed the game very much.

4.5.7 - How improvements could be made in the future?

First of all, making light more important for gameplay would increase the innovativeness of the game and serve as a good demonstrator of the reflection mapping enabler. For example, some enemy waves could be more or less visible depending on light level (prompting players to add external additional light sources or moving to a more strongly lit location), and some waves could even be stealth waves, only detectable by defending forces if pinpointed by the user with a lamp or light probe.

In terms of game dynamics and elements, users recommended the addition of ground elements that have an impact on the gameplay (rivers can slow down attackers momentarily, trees provide them with some cover), enabling the destruction of ground elements and their integration into the game story (blast craters are used as cover by the advancing forces), and allowing for the redeployment of forces between several fixed defence points ("fortresses").

4.5.8 - Did it meet it objectives?

Overall, the Augmented Resistance game was received quite satisfactorily by the experimentation events' attendees, who produced a series of recommendations in several areas. This met the objective of the scenario.

4.5.9 - Summary of outcomes and Conclusion

The Augmented Resistance app demonstrates in a very playful way how augmented reality can be used to make embodied games. It also shows how a light probe allows a more seamless integration between virtual and real content. Test players enjoyed the game, and suggested many improvements at the level of the game play, and at the level of better exploiting the technical capabilities of the setup in the game itself.

4.6 - Scenario "Immersive Control Systems" by running "Dragon Flight"

4.6.1 - Description of tested application

Dragon Flight is a game where the user takes control of a flying dragon. The user's body is mapped one-to-one to the dragon's body. The user will have to move his arms to simulate the flapping wings of the dragon in order to fly and try to pick the coins scattered around the map.

4.6.2 - Test objectives and expected outcomes

Gather feedback and reactions from the users. The data gathered will be very useful to improve the system. We hope that people will find the experience much more immersive than with traditional controller.

4.6.3 - Applied methods and tools for evaluation

Participants were asked to learn to fly the dragon and try to collect the coins around the map. Participants are informally interviewed after playing the game. They were asked about how immersive and fun the experience has been and if they had any ideas for improvement

4.6.4 - Summary of Experimentation in Zurich

4.6.4.1 - Dates (Start/end) of the experiment

The dragon flight playtest took place at DRZ offices the 17^{th} February 2014, where 20 non-expert participants were invited to an informal test of the application and briefly interviewed right after.

4.6.4.2 - Role (involvement) of partners

The dragon flight demo application was entirely developed by DRZ. It uses the Immersive control system SE that was also developed by DRZ and will use the Synchronization GE developed by DFKI and the Leaderboard enabler from ETH. DRZ team took care of setting up the demo and giving support during the event.

4.6.4.3 - Short Report

A booth was prepared were the dragon flight app was deployed. The application setup consists in a Microsoft Kinect placed over a monitor where the game is shown. Participants took turns flying the dragon after a short introduction by the DRZ team. During the informal playtest the participants were watched in general and were briefly interviewed for informal feedback about the experience.

Support was provided by the DRZ team explaining a few tricks on how to fly the dragon properly.

4.6.5 - Summary of outcomes and Conclusion

Feedback was very positive and complimentary. All participants were able to learn to control and fly the dragon. Most of them found the experience very fun and more immersive compared to a traditional controller, such as a game pad. A good portion of the participants complained about the physical effort. Some of them would have liked an even more immersive system, including a head mounted display and a microphone.

This experiment verified the functionality of the Immersive Control System SE and gave us a lot of feedback for improvements.

4.7 - Scenario "Location based virtual reality" by running "Skye Wars VR"

4.7.1 - Description of tested application

The test placed the user wearing an Oculus VR headset within the Skye Wars 3D environment using Augmented Reality Fast Feature Tracking enabler for Tier 2, location based game.

4.7.2 - Test objectives and expected outcomes

The objective was to test the effectiveness of the enabler for use in a wider location based play space, than previously available. The latency of processing between tracking and display of the 3D content is extremely important in reducing potential of nausea and also enjoyment/realism of the game world. It was expected that the low-latency tracking approach would combine well with marker tracking for a robust interactive experience.

4.7.3 - Applied methods and tools for evaluation

The objective analysis of system performance was applied through modelling of time varying sensor data from both IMU (inertial measurement unit) and registered marker pose coordinate streams.

4.7.4 - Summary of Experimentation in Edinburgh

This low cost location based, tier 2, experimentation was performed in Edinburgh with a number of sessions on a small test group to capture the variance of signals over a range of walking, crouching, turning movements.

4.7.5 - What worked well?

The use of hi-frequency low latency orientation sensors, together with low-frequency robust marker tracking was successful in providing a responsive tracking 3D VR experience in a room sized space. The marker tracking integration importantly corrects yaw drift present in many gravitational sensor solutions.

4.7.6 - How improvements could be made in the future?

The fiducial marker tracking could be faster and more robust to lighting conditions with optical spherical point tracking, possibility further enhanced as retro-reflective markers.

4.7.7 - Did it meet it objectives?

The immersive experience resulting was reported enjoyable by participants, and broke ground on allowing immersive interactive walkable spaces of the 3D game environment with low-cost web cams.

4.7.8 - Summary of outcomes and Conclusion

We carried out a quantitative study of the effect of noise and drift on the tracking of a VR headset in a walkable area with modern hardware. Our dual filter approach resulted in low-latency tracking with robustness to yaw-drift.

4.8 - Scenario "Tabletop AR games" by running "AR Travelers"

4.8.1 - Description of tested application

This application is a multi-player game, AR Travelers, in which players have to prevent aliens from invading earth. They do so by interacting with a physical cube that is augmented with CG content, and acts as a dimensional portal in the game.

The evaluation of AR Travelers consists of evaluating the effect of motion blur and latency on the gameplay

4.8.1.1 - Motion blur

Cameras capture an image by exposing the photographic film or digital sensor to photons. Depending on the brightness of the scene being photographed, that is, the number of photons arriving at the film or sensor, the camera settings must be adapted in order to capture a well-exposed image. The camera settings comprise the exposure time (shutter speed for video), the film sensitivity or sensor gain, and the aperture of the lens. For darker scenes, as the sensitivity and the aperture are typically more limited (especially in small form cameras as integrated into mobile devices), the exposure time is often increased. This inevitably leads to blurred images if objects in the scene move relative to the camera. AR applications rely on immersion created through seamlessly mixing the rendered virtual content (usually in the foreground (FG)) with the camera image (usually in the background (BG)). If the FG image does not visually match the BG image, we expected that the immersion might break. Our goal was to test situations where the FG blur matched the BG blur and situations where it did not match. As it is more feasible to add motion blur to the virtual rendered content than to remove it from the camera image, we chose the former method. In this experiment we chose a bright room, lit by daylight through windows and by lamps (see Fig. 1a). In that way the exposure time is short, resulting in images with very little motion blur. Then, we artificially blurred the BG and the FG independently to simulate a longer exposure time and examined the impact on the task performance and the players' perceptions. The amount and direction of the motion blur depends on the camera motion and is calculated as a translation relative to the marker.

4.8.1.2 - Latency

In video see-through AR games and applications running on mobile devices and HMDs the virtual content is rendered onto the camera image. Until the final image is visible on the device's display, processing, synchronization and signal transmissions cause delay and add up to the total system latency. This latency is an important factor for reactive games as well as for augmented reality applications. We used the same game as in the motion blur experiment (Experiment 1) to investigate the influence of latency on the task performance.

4.8.2 - Test objectives and expected outcomes

4.8.2.1 - Motion blur

During the experiment, each participant played 5 rounds, 60 seconds each. In each round, a blur configuration scenario was randomly selected. We presented the game to the player without modification (scenario A), with artificial motion blur only added to the BG (scenario B), and with artificially added blur to both the FG as well as the BG (scenario C), see Table 1. Scenario A is the default case in a bright environment. Scenarios B and C simulate a darker scene when the device's camera switches to a longer exposure time. However, in scenario C there is also artificial FG blur added to match the BG blur, which we expected to create a more realistic mixing of virtual and real images than in scenario B. The player was not informed about the intention of the experiment and thus was not aware that there would be artificial blur added to the game.

Table 4: Configurations scenarios for the motion blur experiment.

	Foreground Blur	Background Blur
Scenario A	Off	Off
Scenario B	Off	On
Scenario C	On	On

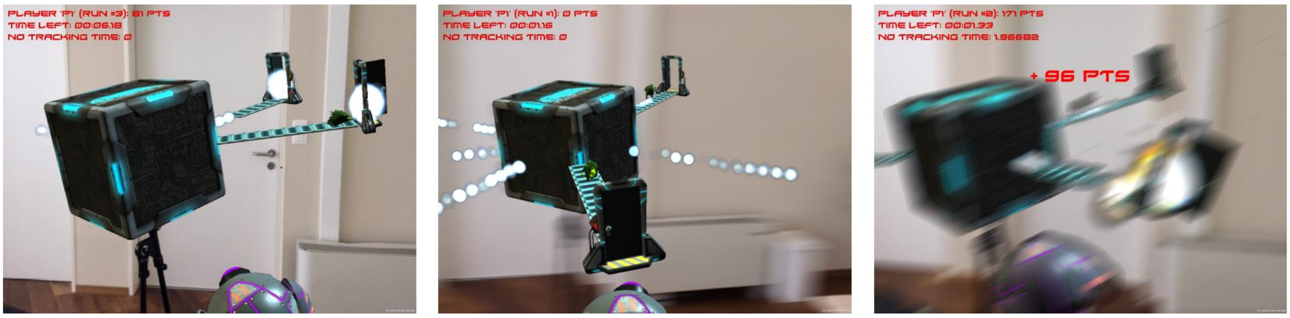


Figure 6: ARTravelers in-game screenshots with different blur configuration scenarios (A, B, C) during the Camera Motion Blur experiment

(a) A: FG blur = Off; BG blur = Off (b) B: FG blur = Off; BG blur = On (c) C: FG blur = On; BG blur = On

Error! Reference source not found. depicts screenshots from the ARTravelers game, modified for this experiment. During the game, the user sees the cube with targets spawning from its center. These targets have to be destroyed by walking to the target and aligning the device with the target and the marker cube. The head-up display shows the collected points, the remaining time, and the time during which the tracking was lost.

We chose to add a relatively strong blur, compared to the naturally occurring blur in dark settings, because we wanted to investigate if the participants would notice the blur at all and if the blur had any influence on task performance.

After each round, the participants were asked to answer the following questions:

- Enjoyment: 'How much did you enjoy this run?'
- Score satisfaction: 'How satisfied are you with your score in this run?'
- Realism: 'How realistic did the game look in this run?'
- Matching: 'How well did the virtual content (foreground) match with the camera image (background)?'

Each question could be answered on a scale from 1 (not at all) to 5 (very much).

4.8.2.2 - Latency

In this experiment, we artificially increase the latency of the frames presented to the players and record again task performance as well as usage experience. We hope to gain insight about the degree to which latency influences these measures in order to judge the importance of reducing sources of latency.

Table 5: Measured native system latency for different settings.

	Lighting	Features	App	Avg latency
Scenario A	Bright	-	iOS camera app	89.9 ms
Scenario B	Normal	High	ARTravelers	101.5 ms
Scenario C	Normal	Low	ARTravelers	97.3 ms
Scenario D	dark	Low	ARTravelers	131.1 ms

Figure 7: Latency measurement setup. The mobile device running ARTravelers is pointed at blinking LEDs. A high-speed camera captures both, the LEDs and the LEDs on the display. The recorded video was analyzed to calculate the delay between the LEDs and the LEDs on the display

4.8.2.3 - Motion blur

The program was modified to record statistical data and send it to a connected server.

4.8.2.4 - Latency

We modified ARTravelers such that it delays all shown frames by N frames to introduce additional artificial latency (AAL), that is, latency additional to the existing native system latency. The application runs at constant 30 fps and therefore every AAL frame adds an additional 33.3 ms latency to the native latency. To verify, we measured the AAL using the same method as in Section 5.1. As expected from the results in Table 2, there is a minimal total latency of about 115 ms. For higher AAL, the native latency decreases to about 50 ms to 60 ms.

4.8.3 **- Summary of Experimentation in Zurich**

For the experiment, the game was played with a single player not requiring network connection. For analysis purposes, we used a local Wi-Fi connection to transfer measurements. The participation was voluntary and the participants could interrupt or abort the experiment at any time.

4.8.3.1 - Motion blur

The participants were in the upper age range of the targeted users. Potentially, the game is also interesting for children. However, for our purposes, we do not expect significantly different results with children.

The first out of five rounds was considered a training round and was omitted in the further analysis to remove a strong influence of the learning effect. This learning effect can clearly be observed in Figure 8. We recorded 12 participants and thus, 48 rounds in total. The experiment took place on July 14, 2014.

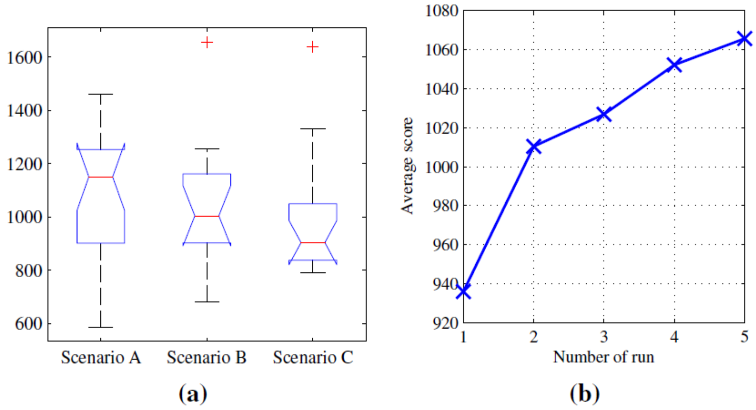


Figure 8

Figure 9: *(a) depicts box plots of scores for each scenario. Visually, a decreasing trend is noticeable from scenario A to B to C. That is, the stronger the total blur, the less the players scored. However, the trend is not statistically significant. (b) shows the average score per round for all players. A learning effect is clearly visible.*

4.8.3.2 - Latency

Similar to the blur experiment, each participant played multiple rounds of ARTravelers with different amounts of AAL and tried to achieve as high a score as possible. The participant was not informed about the intention of the experiment, i.e. that there would be increased latency in the game. First, the participant played two introductory rounds: one without AAL and one with a high (333 ms) AAL. These two introductory rounds were not used for further analysis. Then the participant played another five rounds, each lasting 40 seconds, with AAL bucket randomly assigned to each round.

We intended to sample the lower AAL space denser and defined the buckets as follows:

- Bucket 1: 0 to 1 frames (0 ms to 33 ms) AAL
- Bucket 2: 2 to 4 frames (66 ms to 132 ms) AAL
- Bucket 3: 5 to 9 frames (165 ms to 297 ms) AAL
- Bucket 4: 10 to 16 frames (333 ms to 528 ms) AAL
- Bucket 5: 17 to 35 frames (561 ms to 1155 ms) AAL

Over the five rounds, each player was assigned each bucket once. Inside each bucket a random value was generated for each player. After each round, the participants were asked to answer the following questions:

- Enjoyment: 'How much did you enjoy this run?'
- Score satisfaction: 'How satisfied are you with your score in this run?'
- Responsiveness: 'How responsive did the game feel?'

Each question could be answered on a scale from 1 (not at all) to 5 (very much).

The experiment took place on July 14, 2014.

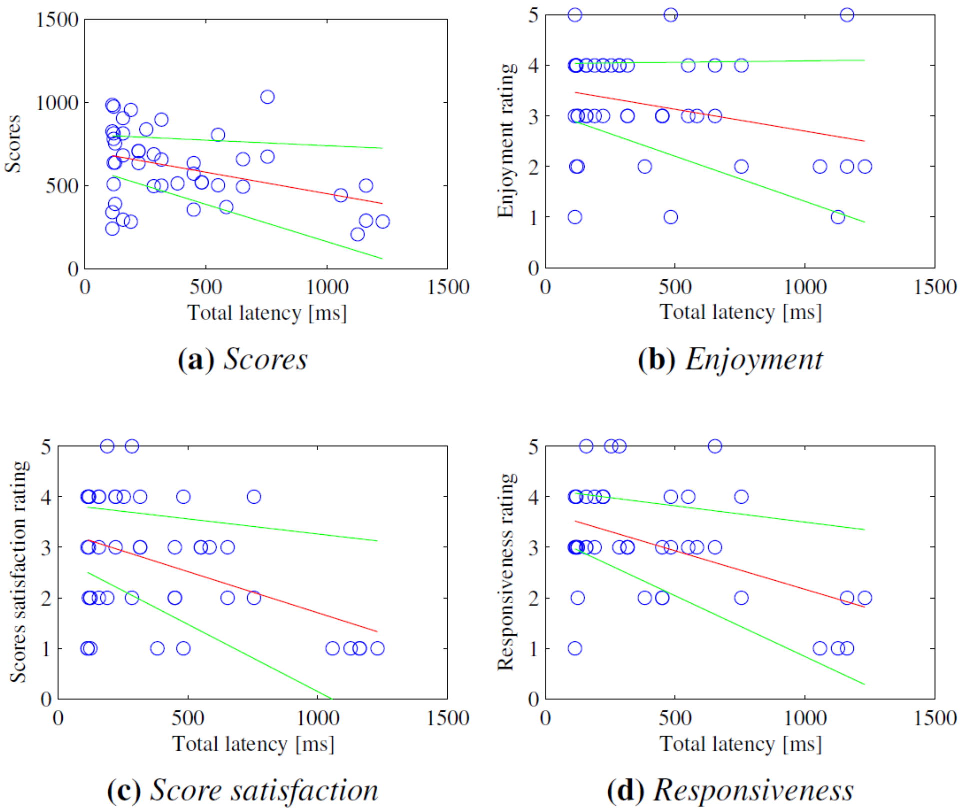


Figure 10: Linear regressions of scores, enjoyment, scores satisfaction and responsiveness over total latency. The red lines depict the fitted linear curves and the green lines define the lower and upper limit of the 95% confidence interval.

The experiment showed a significant (p < 0:01) connection between latency and scores as well as between latency and responsiveness.

4.8.4 - What worked well?

We could successfully run the experiments and extract useful scientific data, and gain an understanding of the importance of the different elements to create a believable experience for players engaged in augmented reality games.

4.8.4.1 - Motion blur

The hypothesis that the blur configuration scenario does not influence the player's performance or experience cannot be discarded (p > 0:05%). This may be due to the fact that there were not enough participants recorded or that all participants experienced the blur scenarios very differently. Some players might have looked over the mobile device to orient themselves in the real world and ran directly to the target's position. Others might have been more confused by the blur and scored less.

4.8.4.2 - Latency

As expected, but as opposed to the blur experiment, the player could perceive a delay in the game and at the same time his or her performance suffered during high latency rounds, see **Error! Reference source not found.**. However, latency had no significant influence (p > 0:05) on the reported enjoyment of the participants. Even in ARTravelers, which requires high concentration but little reaction skills, a negative effect of latency on task performance is apparent.

4.8.5 - Did it meet it objectives?

We could successfully study the effect of motion blur and latency on player experience. This work was published in 'Influence of Animated Reality Mixing Techniques on User Experience', Proceedings of the Seventh International Conference on Motion in Games (Los Angeles, USA, November 6-8, 2014), pp. 125—132.

4.8.6 - Summary of outcomes and Conclusion

Thanks to AR Travelers, we could successfully gain insights on what range of values for motion blur and latency are acceptable for players of AR games. This provides a solid scientific basis to make optimal technical choices for further games.

4.9 - Scenario “Tabletop AR games” by running “Live Inspector”

4.9.1 - Description of tested application

In this section, we analyse the feedback of a Tabletop AR games scenario in a setup where users use AR to program and debug the behaviour of an autonomous mobile robot. This application pushes the concept of user-created content to its limits, and also stresses the demand on the AR hardware and software. In a previous work, we showed that students using the educational robot Thymio (http://thymio.org) and its visual programming environment (VPL) were able to learn the important computer-science concept of event handling. The tested application, call “live inspector”, extends that work by integrating augmented reality (AR) into the activities using the Visual Agent Design SE. Students use a tablet that displays in real time the event executed on the robot, providing a program-tracing tool. The event is overlaid on the tablet over the image from a camera, which shows the location of the robot when the event was executed. In addition, visual feedback was implemented in the software.

4.9.2 - Test objectives and expected outcomes

The goal of this experiment was to answer the question of whether the use of augmented reality through the live inspector improved the learning of computer science skills during a short-duration workshop. We run two workshops in parallel in two different rooms. There were two independent variables in this experiment:
• Augmented reality: AR was used by the students in room 1, while in room 2 VPL was used without AR
• Visual feedback: The first session in each room used visual feedback from VPL and the AR system, whereas visual feedback was not used during the second session.
The expected outcomes were a validation of the effect of augmented reality on the learning of computer science and the comparison with an alternative method: visual feedback in the program editor.

4.9.3 - Applied methods and tools for evaluation

For the current research we used a novel type of questionnaire based upon video clips (details can be found here: http://thymio.org/en:thymiopaper-vpl-iticse2015). There were eight questions, four each of the following types:

- The student is shown a video of the behaviour of a robot and then asked to select a program (one out of four) that causes the behaviour.
- The student is shown a program and four videos and then is asked to select the video that demonstrates the behaviour of the robot running the program.

The questionnaire was constructed using the Forms facility of Google Drive. In Forms, you can include videos by uploading the videos to YouTube and providing the URL.

We also collected usage data by instrumenting the VPL editor.

4.9.4 - Summary of Experimentation in Zurich

The workshops consisted of 14 sessions of 75 minutes. Two sessions were run in parallel in two rooms. The workshops took place in Lugano, Switzerland on October 16-17 2014. The first 15 minutes of each 75-minute session were devoted to introducing the robot and its basic built-in behaviours. During the next 15 minutes, the students learned about the VPL environment and this was followed by 30 minutes devoted to solving increasingly challenging tasks. During the final 15 minutes the students answered the video questionnaire.

The experimenters were four teaching assistants, two per room, who were students at USI (Università della Svizzera italiana). The same two assistants were always paired together. After every two sessions, the assistants exchanged rooms to prevent bias due to a specific pair of assistants. The experimentation was oversight by a FIContent2 member. There were 10-18 high-school students per session from high schools in the Swiss canton of Ticino. The median age of the students was 16 (low/high quartiles: 16/17).

We used a mix of Android and iOS devices, with screens ranging from 7 to 10 inches. The students were free to opt-out of the study. For students who agreed to data collection, we collected usage data from the VPL editor: addition and deletion of blocks, change of parameters, and clicks on buttons. For the AR system, we collected usage data from the tablet: its position in 3D, the position of the robot in 2D, and the state of the application. In addition, we asked these students to answer the video questionnaire.

4.9.5 - How improvements could be made in the future?

All students could successfully program the robot, and make little mistakes in the questionnaire (15 % in average).

To compare the treatment vs. the control groups, for every condition we counted the number of mistakes for every participant. The following table shows the mean mistake count and the p-value of Pearson's chi-square test its histograms, for the null hypothesis of no effect. We used Laplace smoothing (adding 1 to each bin) to apply the chi-square test even when the control group has 0 entries for a given mistake count:

	Treatment	**Control**	**p-value**
Feedback	0.81 (n = 47)	1.74 (n = 34)	0.003
Augmented reality	1.00 (n = 41)	1.40 (n = 40)	0.10

We see that both AR and feedback reduces the error rate, but only feedback leads to a statistically significant result. Additional qualitative observations of how to improve AR were collected (see next section).

4.9.6 - How improvements could be made in the future?

The AR hardware and software needs to be made more robust and easy to use, for AR to be suitable for activities in which it is not the primary focus. In addition, when AR is to be used in education, learning materials specifically designed for AR must be developed. In particular:

- The groups with AR required intensive support because AR significantly increases the complexity of the setup.
- Some students found using the tablet as a live inspector to be unintuitive.
- Several students tended to keep the tablet too close to the robot and therefore the tablet did not see the ground image.
- The use of the tablet was not uniform: some students did not use the tablet at all, while others seemed lost in contemplating reality through the tablet.
- The students did not always realize whether the tablet was tracking the ground or not.
- The software did not work uniformly well on different devices, especially those with different screen sizes. It is actually challenging to write an application that fits devices spanning a large variety of formats.
- Energy consumption was a problem. Scenarios using AR should try to use it sparsely, to save battery life.
- Some tablets have poor focusing abilities and sometimes stayed out of focus for several minutes.
- The AR system sometimes lost track of the robot. Even current AR technologies would need improvement to handle such challenging scenarios.

4.9.7 - Did it meet the objectives?

Yes, we could validate the functioning of the Visual Agent Design SE and contribute to the state of the part of computer-science education. As a proof, a research paper on this study, entitled "Enhancing Robot Programming With Visual Feedback and Augmented Reality", was accepted for publication in ITiCSE 2015, a major international conference in innovation in computer science education (30 % acceptance rate).

4.9.8 - Summary of outcomes and Conclusion

We carried out a quantitative study of the effect of visual feedback and augmented reality on the learning of a computer science concept using a mobile robot. Visual feedback had a significant positive effect on some questions, while AR had a positive effect but the improvement was not significant, although it did modify significantly the user behaviour.

4.10 - Scenario "City-Wide Economic Game" by running "Gnome Trader"

4.10.1 - Description of tested application

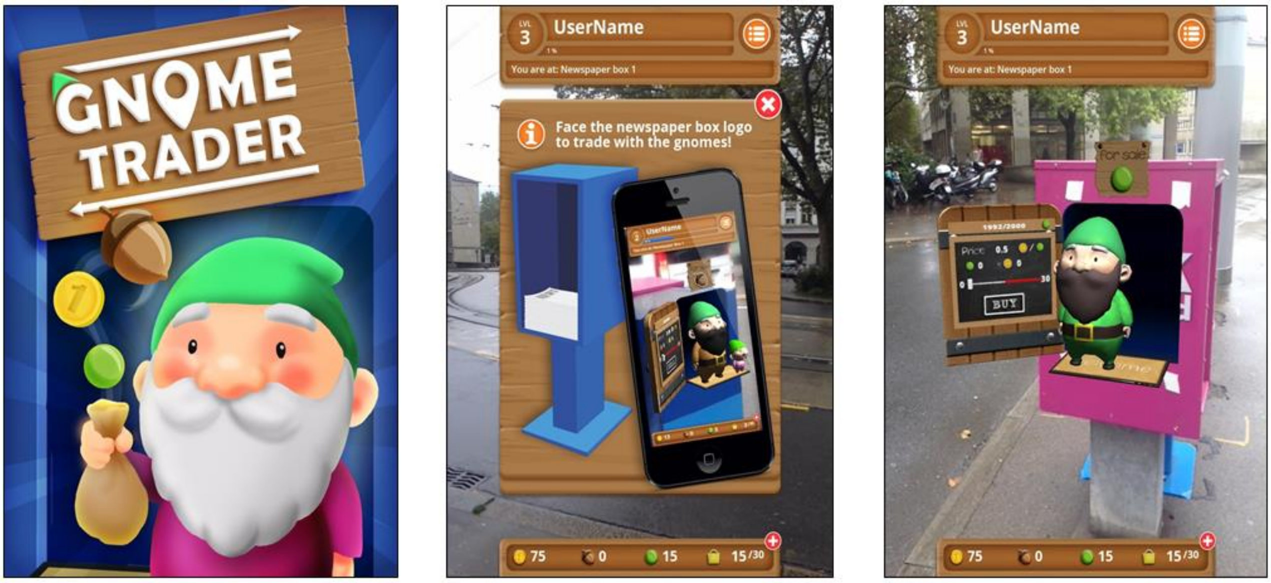


Figure 11: Gnome Trader game application

Gnome Trader is a city-wide location aware Augmented Reality resource trading game. Resources can be bought from and sold to Gnomes hidden in recurring locations all around the city. Gardener gnomes produce food and sell it to the player at a dynamic price. Gnome families are waiting for the player to bring them food. Additionally, the Gnomes change their resource prices according to the amount of resources in their storage.

4.10.2 - Test objectives and expected outcomes

The goals of the user trial were:

- To test the field deployment of the Gnome Trader game application, with a range of Android devices and operating systems, and gauge the operation of its Enablers.
- To validate the attractiveness of the game for a sample of experienced and non-experienced gamers, and obtain insights on how to refine the game.
- To obtain sufficient user-generated game session data to improve the game economic mechanisms and cheat detection modules.

4.10.3 - Applied methods and tools for evaluation

The evaluation framework of this experiment was based on a qualitative evaluation with an online questionnaire with SUS and UX items, and a quantitative evaluation based on backend statistics. This allowed for a situation in which the qualitative feedback could be used to contextualize and understand the user behaviour that was observed through the quantitative analytics.

Summary of Experimentation in Barcelona.

The technical infrastructure for the experiment consisted in the user-contributed devices and internet connections. A switch was provided at the testing venue for users to access the web service for the first time. The web-based component of the application was initially deployed at FI-LAB, but technical problems made it advisable to redeploy the enablers at a more stable location. This caused serious delays and complications in the conduction of the experimentation, including the rescheduling of the competition. For this reason, the experiment infrastructure provided by DRZ was used to deploy the enablers on which the game runs.

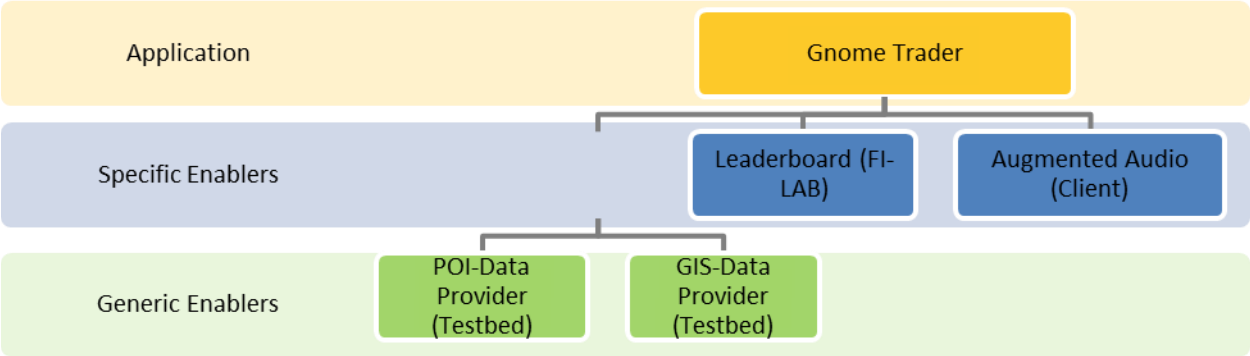


Figure 12

The competition, which was initially planned for January 23rd, had to be rescheduled to a later date due to technical problems with FIWARE deployments.

A special-purpose application was built by ETHZ to test the selected marker image in situ. After lab and field validation, this marker image (see Figure 12 above) was approved for usage in the Barcelona version of the Gnome Trader application.

A set of 10 game locations was created around the experiment venue, throughout the neighbourhood of Sant Andreu. Six were naturally-occurring images, already available at public transportation stations of the neighbourhood, and four were temporary locations created for the purpose of the competition.

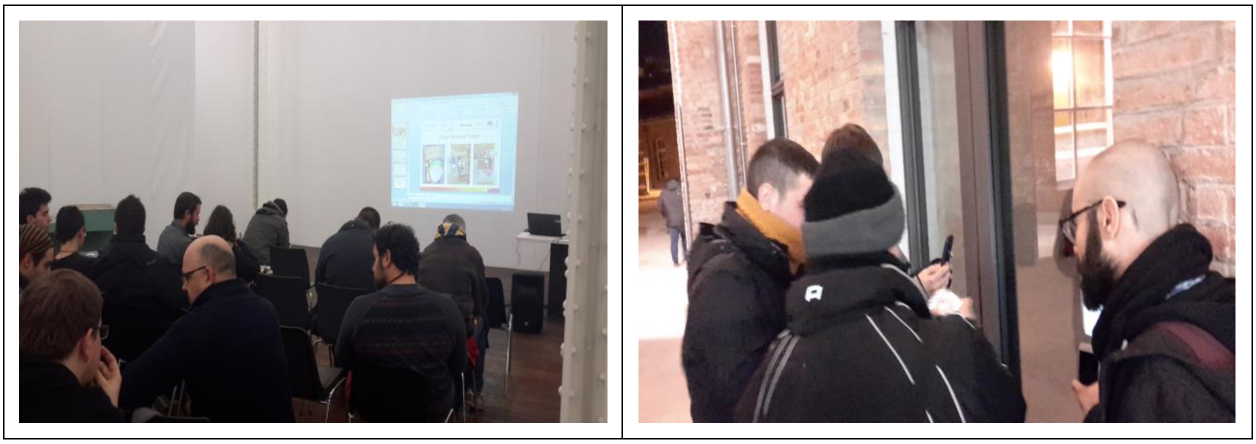


Figure 9 Participants taking part in Gnome Trader competition, 6th February 2015

When all users had created their game accounts, the competitive game session started. Users constantly tried to outsmart each other, trying to get the best prices at every location. After approximately one hour and a half of gameplay, the users returned at the experiment venue, and filled in a feedback questionnaire. The winners were determined automatically by the Leaderboard SE, which aggregated the total scores of each participant.

4.10.4 - What worked well?

The analysis of the responses to the items of the feedback questionnaires disclosed that the participating sample of users perceived the application in a generally positive view. As a means to generate a single figure that would encapsulate the user's satisfaction with the service, the recommender score was used. The average results for the recommender score were very positive, standing at 73.5%.
Additionally, competition participants were also asked about how often they thought they would like to play this game in their everyday life (see Figure 13 below).

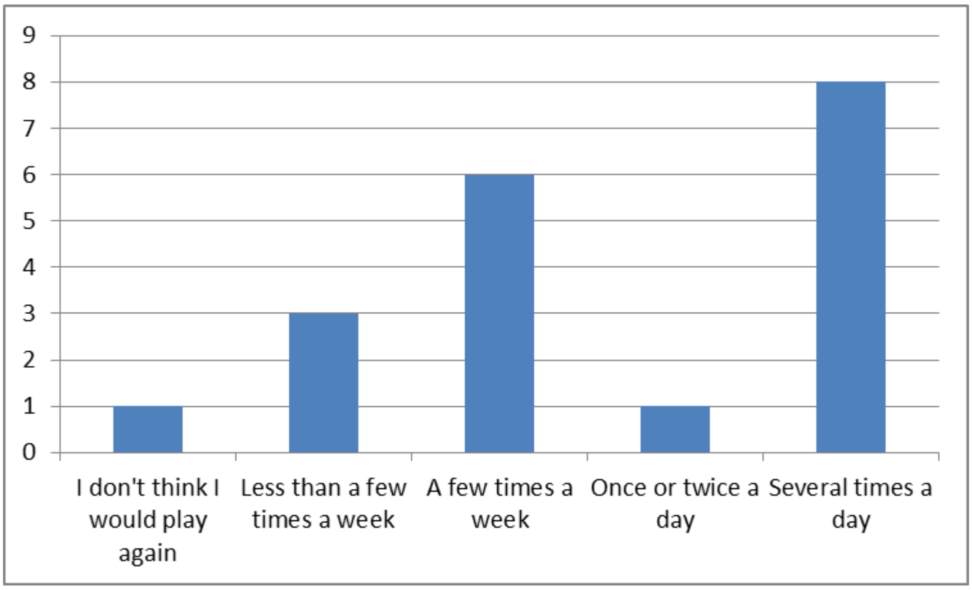


Figure 13: Gnome Trader user trials: responses to question "how often would you like to play?"

The results gathered from the responses to this questionnaire item were quite positive, with an overwhelming 95% of users declaring being interested in playing the game again. The pattern of responses clearly defines two groups of users, those who think that would play occasionally (45% of the sample), and those who would be willing to play several times a day (40% of tester group).

Other aspects of the user experience that were evaluated with the final questionnaire were the perceived interest, ease of use, graphics, innovativeness, rules of the game, and use of augmented reality and audio on the Gnome Trader game. On the whole, the user evaluation of these dimensions of the gameplay experience was fairly positive. The interest, rules of the game and use of augmented reality of the application were particularly praised.

4.10.5 - How improvements could be made in the future?

A majority of users (40%) had no problems at all using their period of usage of the Gnome Trader application, as can be seen in Figure 14 below. On the other hand, 35% of testers had at some point some kind of technical problem, which prevented them from taking full advantage of the application during the competition. Most of these problems had to do either with the buy/sell sliders being difficult to manipulate on the very small screens of lower-end devices, or with problems with the marker tracking in very dark lighting conditions. The remainder quarter (25%) of users experienced some small degree of technical difficulty during the experiment, mostly related to the layout of the application's graphical interface in different screen sizes, but were not at any moment unable to participate in the competition.

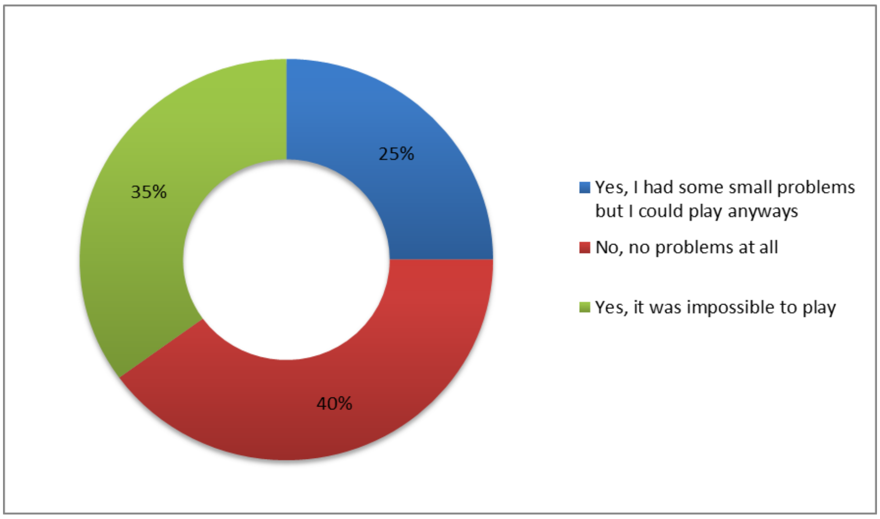


Figure 14: Types of technical problems across devices in Gnome Trader user trials

These technical issues had an impact on the application's perceived ease of use, which was given a below average rating.

4.10.6 - Did it meet the objective?

The experimentation successfully demonstrated that the enablers packed into the game were able to handle and process the user activity, efficiently and without critical problems, in a real-life usage scenario. The game's economic engine adjusted as expected to changing levels of supply and demand, triggered by the users' buying and selling activity. These changes led to a rich and competitive gameplay experience, and to the emergence of complex agent strategizing to maximize profits, amass points and win the game. Although the wide range of hardware qualities and smartphone models in the set of user-contributed devices led to a higher than anticipated number of technical issues, there was generally a high level of user engagement in the gameplay throughout the user trials. The analysis of the participants' responses also revealed a high level of interest in the game concept, and fairly good ratings in standard measures of usability, satisfaction and user experience.

Other aspects of the user experience that were evaluated with the final questionnaire were the perceived interest, ease of use, graphics, innovativeness, rules of the game, and use of augmented reality and audio on the Gnome Trader game (see Figure 15 below).

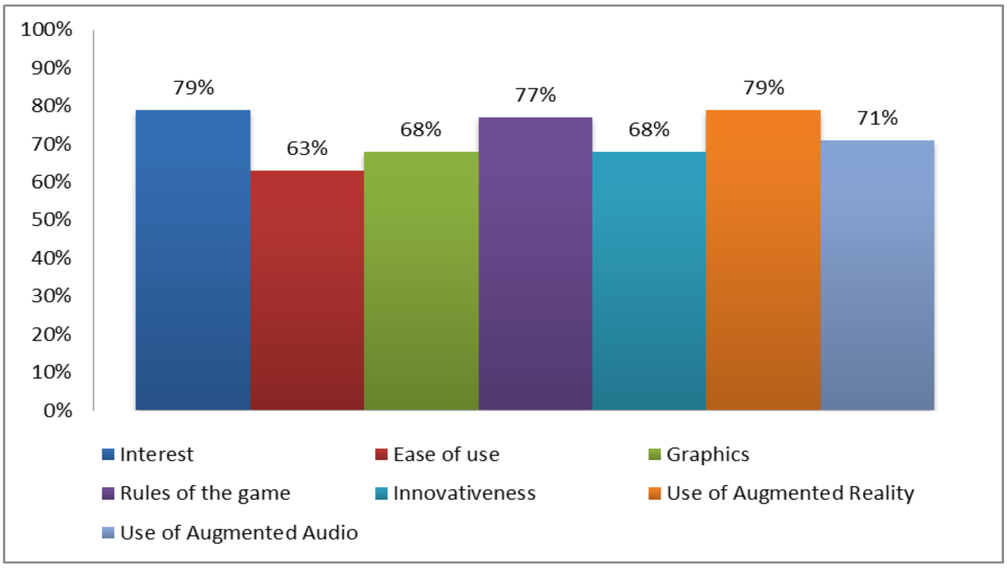


Figure 15: User rating of dimensions of user experience with Gnome Trader

On a less positive note, the application's ease of use was given a below average rating, due to the technical issues laid down in detail at the beginning of this section.

All objectives were met.

4.10.7 - Summary of outcomes and Conclusion

The main recommendations that arise from the experimentation in this scenario in the Barcelona site are:

1. According to the qualitative and quantitative data generated at the experiment, the Gnome Trader game application has reached a satisfactory level of technical performance and user evaluation. The qualitative data show that the users perceived in an especially favourable way the application's interest, usage of augmented reality to enrich the game, and the exciting dynamics of the gameplay. On the other side, the quantitative data attest to a high level of user engagement, and well-functioning game dynamics in real-life conditions.

2. Although acceptable in the current version of the application, the game's usability could be improved if certain issues were addressed. In particular, ways to improve the recognition of the marker in less-than-optimal lighting conditions (i.e. at night or in dimly lit areas) should be explored. Also, the game's power consumption seems to be on the high side, which could be a hindrance to playing the game. Battery-saving improvements would do much to enhance the practical attractiveness of the game in the eyes of prospective users.

3. The public release of the game on Google Play is recommended, as it would expose the game to a larger number of users, and would allow to validate the conclusions of this user trial with a larger and more diverse sample of users.

4.11 - Scenario "City Wide Augmented Reality Strategy Game" by running "Outdoor Tower Defense"

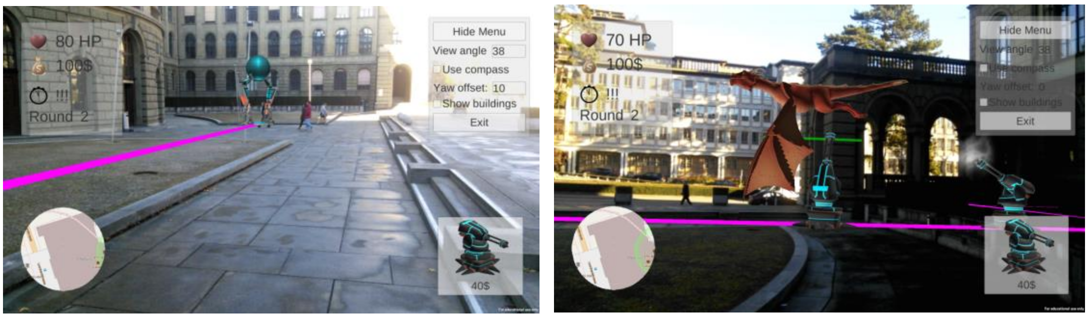


Figure 16: *Screenshots from city-wide AR Tower Defense game*

4.11.1 - Description of tested application

The game is an outdoor city-wide tower defense game. It can be played by multiple players that have the common goal of defending the base.

Unfortunately, SLAM based tracking technologies were not yet available to create this game. In order to explore the game play of such games and test playability with tracking based on other sensors, this game was developed.

4.11.2 - Test objectives and expected outcomes

The accuracy of the tracking was tested by comparing consistency across several devices. The back projected locations of 3D structures and absolute directions were compared as well as repeatability tests were performed.

As this type of game is novel, we were also interested in any insights regarding the game play. We hoped to gain a better understanding of the needs of the players.

4.11.3 - Applied methods and tools for evaluation

Several subjects were asked to play the game and their feedback was collected. Some users were also observed to study their reactions and behaviour.

In addition, sensor data was logged during experiments to get an idea on tracking accuracy in the field tests, i.e. outside lab conditions.

4.11.4 - Summary of Experimentation in Zurich

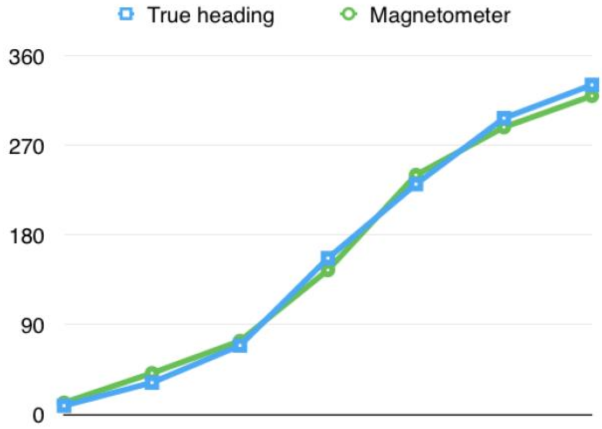


Figure 17: *Angular differences measured during experimentation. While the deviations look small, they still have a large influence.*

The measured position (derived from GPS coordinates) jumped sometimes by several meters. The calibration of the magnetometer was not perfect and lead to a strong (15 degree) discrepancy between devices. While such tracking problems were expected, we could gain insight how much this affects game play.

4.11.5 - What worked well?

At latest when playing the game, high tracking requirements for such games became obvious. Even though the used tracking method made the game hard to play, the testers still enjoyed the novel game concept. One user said in the feedback form: “Once these positioning issues are resolved, it’s gonna be a big thing”.
This experiment made it very clear that camera based tracking information needs to be included. In the simplest case, a 2D stabilization would help. In the best case, a SLAM method would replace the tracking part.

4.11.6 - Did it meet it objectives?

Yes: while the tracking does not work as well as we hoped, we much better understand the requirements now.
The tests helped a lot in understanding requirements of technologies and better understanding game concepts for city wide AR games. Besides the mentioned tracking or positioning of the device, we experienced the connectivity difficulties via the 3G network.

4.12 - Scenario “Creating slam-based AR game using web creator tool” by running “ARvatar”

4.12.1 - Description of tested application

The ARvatar game demonstrates how the web tool ARTool (**Error! Reference source not found.**) can be used to create engaging AR applications. The ARTool aims to dramatically simplify the development and deployment of AR applications by utilising pre-defined templates, components and modules enabling the creation of cross-platform applications with minimal or no programming effort required.

The ARvatar game (Figure 19) is based on marker-less or marker AR tracking where the AR content is superimposed on a global plane detected (e.g. on a street pavement or on a table top) or 2D image marker respectively. Each user chooses its own ARvatar from a set of available 3D models (currently only one model is available). The ARvatar's appearance can be influenced by many factors but initially it will be dependent on the meteorological and air quality parameters determined by the ekoNET service (for now only air temperature and concentration of CO2). This is an IoT device connected to the ARTool platform providing real-time measurements on different meteorological parameters such as temperature, air pressure, humidity and air quality. Depending on the temperature and air quality, the ARvatar’s appearance will be modified accordingly. The users can view the ARvatar and share the view with other users via Facebook. The aim of the game is to score as many points as possible by trying to guess the true values of measured temperature and concentration of CO2 based on the appearance of the ARvatar. The score of the user is logged on the server where the leader board is kept. The measured values of temperature and CO2 are obtained from the device selected from the list enabling users to become aware of the meteorological and air quality parameters at different locations.

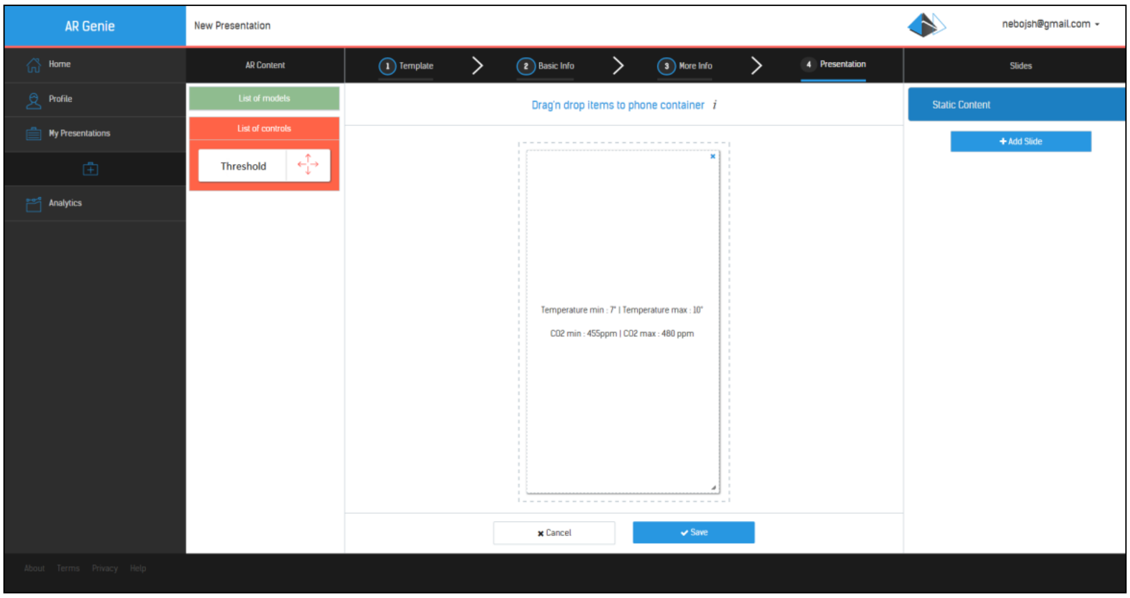
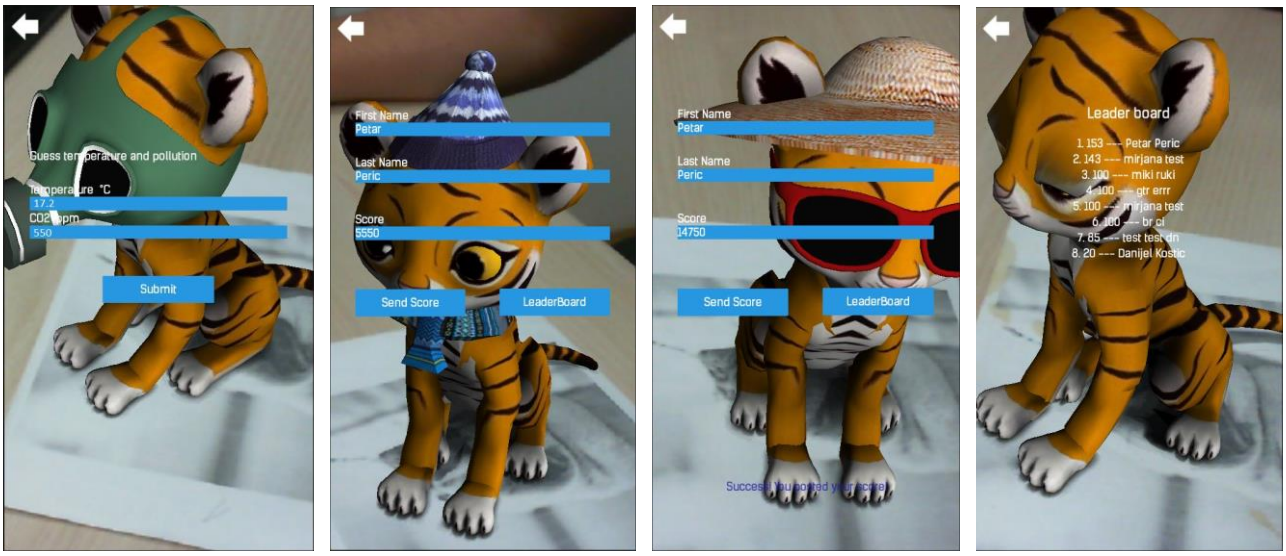


Figure 18: ARTool portal



a) b) c) d)

Figure 19: Mobile application look and feel.

4.12.2 - Test objectives and expected outcomes

The aim of the experiment is to determine an initial opinion of end users of the initial release of ARTool platform and ARvatar game created with the ARTool on the following topics and questions:

- Does ARTool simplify the development of AR applications?
- Is the usage of ARTool intuitive enough?
- Does ARTool provide sufficient amount of features for the development of AR applications?
- Is it possible to use ARTool by the non-technical persons?
- What is the performance of AR applications on the mobile platform being used?
- Is ARvatar game educational and entertaining?
- Is ARvatar game simple and intuitive to use?
- Does ARvatar game promote environmental issues?

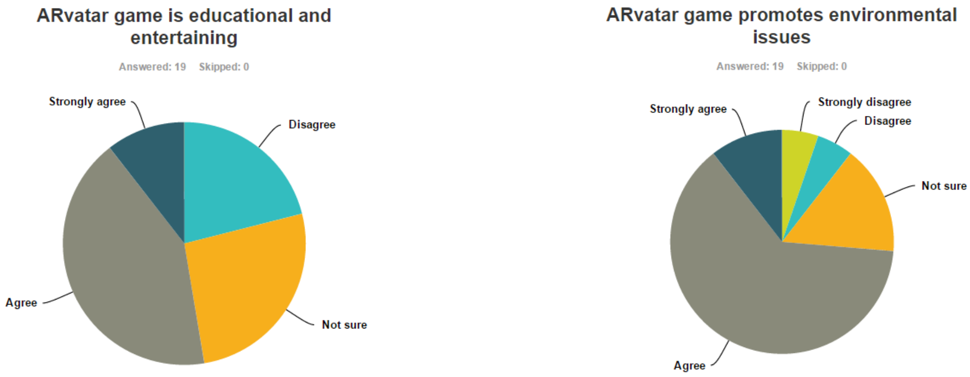
- Does ARvatar game provide a basis for further development within the community?
- Does ARvatar game perform well on the mobile platform being used?

The expected outcome is to receive valuable feedback on the ARTool and ARvatar game in terms of the functionality and other aspects as mentioned above in order to perform reactive actions if required so that the products and enablers can be appropriately improved in the subsequent releases.

4.12.3 - Applied methods and tools for evaluation

The user opinion is obtained using the classical survey methods using the web-based tool SurveyMonkey where the users are asked to provide answers to five questions related to the usage of ARTool and another five questions related to the ARvatar game. The users are first asked to create simple application using ARTool and test it on mobile (it could be ARvatar game, photo shooting or treasure hunt game). Subsequently, the users are then asked to play the ARvatar game. Following these two tests the users are then presented with the web-based survey to provide answers to the questions and potentially the comments on how to improve both applications.

4.12.4 - Summary of Experimentation in Novi Sad, Serbia



Answer Choices	Responses	
Strongly disagree	0.00%	0
Disagree	21.05%	4
Not sure	26.32%	5
Agree	42.11%	8
Strongly agree	10.53%	2
Total		19

Answer Choices	Responses	
Strongly disagree	5.26%	1
Disagree	5.26%	1
Not sure	15.79%	3
Agree	63.16%	12
Strongly agree	10.53%	2
Total		19

Figure 20: Collected responses from the users

From the collected responses (Figure 20) it is possible to conclude that the initial version of ARTool shows positive trend in all five categories covered by the five questions presented in the survey, namely simplicity for AR development, usability, available features and usage simplicity. Received feedback and associated comments and suggestions will be used to improve the ARTool in the following release.

In terms of the ARvatar game, the results are slightly worse than of those of the ARTool. The survey indicates that more explanations are needed in order to present the main concept of the game. Short story will be provided and a simple tutorial in order to address this issue. The main indicator, which is related to the promotion of the environmental issues, shows that the participants agree that ARvarar game achieves this.

ARTool and ARvatar worked well on different web browsers and mobile platforms. Furthermore, all participants could build the game themselves and positive feedback is received on the overall idea both in terms of web-based ARTool functionality and resulting ARvatar game.

4.12.5 - How improvements could be made in the future?

Main comments indicate that more tutorials for the ARTool are required and more templates need to be available in the future. This is already incorporated in the ARTool roadmap. Additionally, some technical issues have been identified and these will be addressed in the next release. Additionally, further work is required in the field of ARvatar gameplay and stability on different mobile platforms.

4.12.6 - Did it meet its objective?

The objective set were met in terms of obtaining the valuable feedback from the initial set of users.

4.12.7 - Summary of outcomes and Conclusion

The experiment has also successfully served as a test to determine how well the ARTool platform and ARvatar game will be performing with the variety of different users, web browsers and mobile platforms.

4.13 - Scenario “City-Wide Scavenger Hunt Game” by running “Treasure Hunt”

4.13.1 - Description of tested application

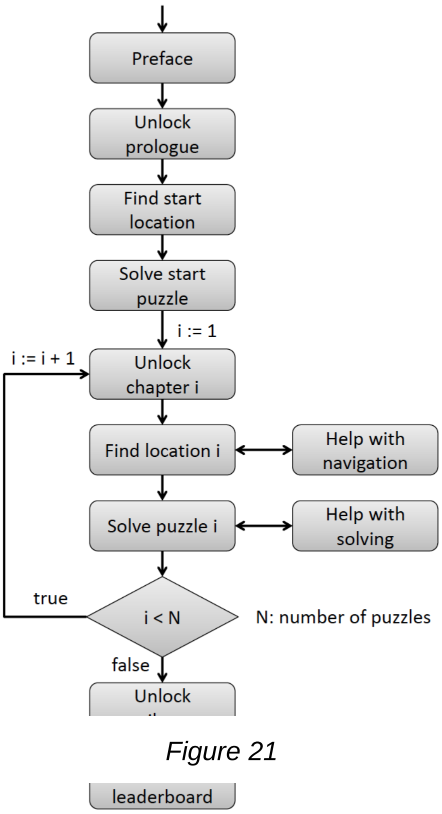


Figure 21

Treasure Hunt Zurich is a city-wide Augmented Reality scavenger hunt game that is played in the city of Zurich, Switzerland. The player is sent on a quest that leads through the city to find a hidden treasure. The locations are found with the help of text descriptions and clues. At several locations throughout the quest Augmented Reality puzzles have to be solved in order to continue. A fully functional city-wide AR treasure hunt game was developed together with a framework, which assists in building similar games. It was implemented using the Unity game engine and Vuforia. The game resembles a traditional scavenger hunt where players sequentially navigate to several locations by following hints and solve a puzzle at each location, as outlined in a game flow diagram in Figure 21. AR is used to augment building facades, which act as image markers, with virtual puzzles on the device’s screen. Solving a puzzle unlocks the next hint to find the next location.

The player’s position is tracked using GPS. When the player arrives at a puzzle location and enables the camera, the position and orientation is estimated to a very high degree by Vuforia’s state-of-the-art visual camera pose estimation algorithms.
The player’s path, the total time needed to complete the quests, and other performance metrics are during gameplay. After finishing the game a high-score is calculated and submitted to a leaderboard, thereby allowing players to compare their performance to each other.

The game is split into 9 levels containing 6 unique puzzles. Some of the puzzles are shown in the Figure 23 to Figure 24 below. A beta testing session was held to evaluate the playability of the game.

4.13.2 - Test objectives and expected outcomes

The objective of the beta testing session was to identify bugs and problems as well as potential loopholes and difficulty issues in the game.

4.13.3 - Applied methods and tools for evaluation

The participants were asked to complete the game and then fill out a questionnaire.

Figure 22: The Music-Puzzle

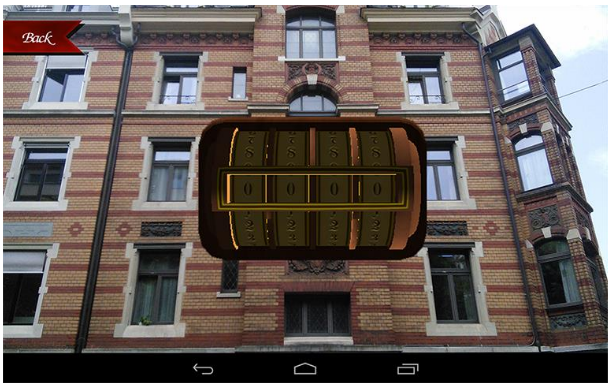


Figure 23: The Combination-Lock-Puzzle

In addition to the questionnaire, performance metrics were automatically recorded for each user during playtime.

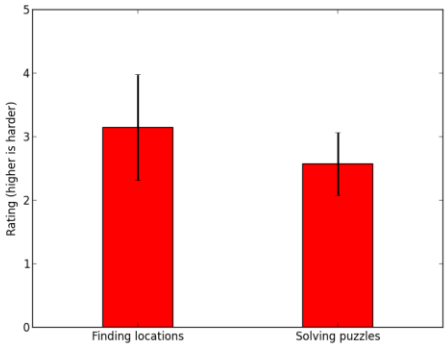


Figure 24: Difficulty metrics

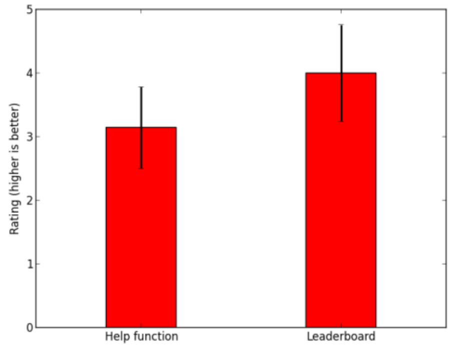


Figure 25: Functionality metrics

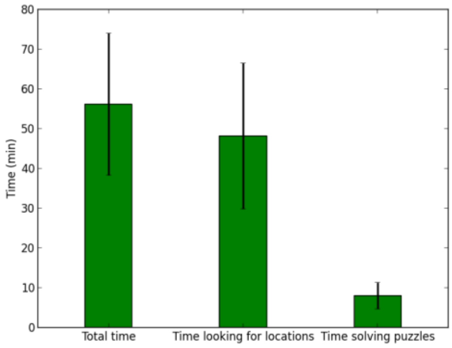


Figure 26: Time metrics

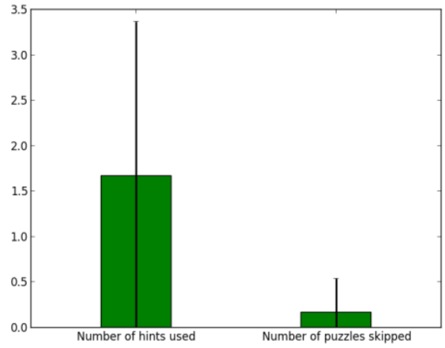


Figure 27: Use of help function metrics

4.13.4 - Summary of Experimentation in Zurich

Two beta testing sessions were held in downtown Zurich on different days with 4 participants in the first and 6 participants in the second session. There were similar lighting conditions during both sessions. Both sessions took place in the early afternoon.

In total 10 beta testers with an age of 22-59 (8 male, 2 female), who were not familiar with the game, participated in the beta testing session. Some of participants played the game on their own while others were observed by the developers to get a better idea of how they played different parts of the game. Each session took about 2 hours. The experiment took place on August 6, 2014.

4.13.5 - What worked well?

The playtesting was a great success and provided useful feedback about possible problems and ideas for improvement. Analyzing the questionnaire we identified the following issues:

Problems
Most issues related to finishing the game resulted from unclear descriptions to find the locations. All but one player could complete the game nonetheless. All texts in question were reviewed and made slightly more precise. Another common problem was the inaccuracy of the compass, which is a limitation of the hardware. To improve the situation, the descriptions were changed to have more indications like "left" or "right" instead of cardinal directions only. Last, the building recognition is lacking reliability at times, which is mostly due to poor lighting conditions. Players playing the game late in the afternoon were limited.

Loopholes
One player found out that the application can be quit to stop the timer when walking from one location to the other and thus cheating by submitting an impossible short time-to-finish-the-game value. However, as the location descriptions are long and hard to memorize, this loophole was not addressed in the next release.

Difficulty
The puzzles were generally considered easy. Only the last puzzle was hard to solve for some players. On the other hand, finding the locations was a challenge for most players. The average beta testing player needed around one hour to solve the whole game whereas a player who knew the locations beforehand could complete the game in less than half an hour.

4.13.6 - What improvements could be made in the future?

The beta tester presented the following ideas for further improving the game:

- More puzzles should be added to create a longer experience.
- A difficulty setting would enable children as well as scavenger hunt experts to play at the same time.
- The current story is simple and could be expanded.
- Other markers than building facades could be used and hence other types of puzzles could be created.

All objectives were met.

4.13.7 - Summary of outcomes and Conclusion

Throughout this project a city-wide AR treasure hunt game for the Android and iOS was developed to show how AR technology can enhance the player experience. The game is fully functional and can be played in Zurich. The underlying implementation of the game can serve as a framework to build similar other games or extend the existing version with more chapters or another story.

Feedback about the game was collected in a formal beta testing session with 10 users. Several extensions and applications should be considered:

Expanded game
The game could easily be extended to be longer or have other, maybe more challenging puzzles.

Multiplayer features
Currently, the only multiplayer aspect in the game is the leaderboard. The game could be modified to allow for multiple players that all compete simultaneously. They could start from different locations and run into each other during the game. Cooperative, adversarial, and manipulative interactions are imaginable. This would most likely require a constant Internet connection during the game and a server-client architecture.

Tourism

The game could include and display tourism information like maps, videos, or other media about the city while the game is played. It would showcase historic landmarks and other tourism attractions and educate the user.

Museums

The game could be adapted to be played inside a museum. The controlled lighting conditions and the fact that the used recognition software is designed for images, using paintings instead of buildings could even increase playability as there would be fewer image tracking issues.Two tests (experiments) have been carried out in order to identify how the end-users are satisfied with the ARTool platform and ARvatar serious game. It has been concluded that the first release provides satisfactory end-user experience and achieves the goals that have been set. Certain functional, user friendliness and performance related issues have been identified which will be corrected in the following releases.

CONCLUSION

In summary this document shows that WP4 have shown that since the very early days of research, development and experimentation it has produced a wide range of technologies that is of significant value to the FICONTENT. It also shows that there is opportunity to add value to the concepts and technology that is now available.

REFERENCES

[1] "Introduction to User Analytics" by Anders Drachen, Alessandro Canossa and Magy Seif El-Nasr, Game Developer Magazine

[2] (August 2012) http://www.mobyaffiliates.com/blog/the-best-mobile-site-and-app-analytics-tools/

[3] (October 2013) http://www.mobyaffiliates.com/blog/list-mobile-app-analytics-tools/

[4] (November 2014) "The Mobile 2D Game Engine Popularity Index" by Steffen Itterheim

[5] http://www.projectskye.ch/

end of the document