CR-PLAY

"Capture-Reconstruct-Play:
an innovative mixed pipeline for
videogames development"

Grant Agreement
ICT-611089-CR-PLAY

Start Date	End Date
01/11/2013	31/10/2016

Deliverable 5.1

Evaluation Plan

Document Information

Deliverable number:	5.1
Deliverable title:	Evaluation Plan
Deliverable due date:	30/04/2014
Actual date of delivery:	30/04/2014
Main author(s):	Christos Fidas, Melpomeni Masoura, Nikolaos Avouris (UPAT)
Main contributor(s):	Ivan Orvieto (TL)
Version:	1.0

Versions Information

Version	Date	Description of changes
0.1	17/03/2014	First draft of the deliverable
0.2	24/03/2014	Containing first ideas for the CR-PLAY evaluation
0.3	01/04/2014	Draft deliverable
0.4	16/04/2014	Pre-final
1.0	29/04/2014	Final

Dissemination Level

PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Deliverable Nature

R	Report	**X**
P	Prototype	
D	Demonstrator	
O	Other	

CR-PLAY Project Information

The CR-PLAY project is funded by the European Commission, Directorate General for Communications Networks, Content and Technology, under the FP7-ICT programme.

The CR-PLAY Consortium consists of:

Participant Number	Participant Organisation Name	Participant Short Name	Country
	Coordinator		
1	Testaluna S.R.L.	TL	Italy
	Other Beneficiaries		
2	Institut National de Recherche en Informatique et en Automatique	INRIA	France
3	University College London	UCL	UK
4	Technische Universitaet Darmstadt	TUD	Germany
5	Miniclip UK Limited	MC	UK
6	University of Patras	UPAT	Greece
7	Cursor Oy	CUR	Finland

Summary

This is the first deliverable from **Work Package 5: "Formative and Summative Evaluation"**. The leader of this work package is **UPAT**, with involvement of the following partners: **TL**, **INRIA**, **CUR**, **UCL**, **TUD** and **MC**. The objective of this work package is to develop a concrete usability evaluation plan embedded in the iterative design and implementation of CR-PLAY's pipeline for videogame development. The purpose of this work package is to evaluate CR-PLAY's method of videogame development vs. traditional reference methods in the frame of formative and summative evaluation studies, aiming to provide evidence on improved efficiency with regards to time and resources used, taking into account the quality of produced assets.

In this context, this deliverable "**D5.1. Evaluation Plan**" describes in detail the evaluation methodology that is planned to be adopted in CR-PLAY. Given that the content creation of videogames is a multidisciplinary activity, it is necessary to evaluate the CR-PLAY approach from several perspectives and diverse users' points of view. From an end-users perspective the main evaluation factors are: a) the CR-PLAY tools (Capture, Reconstruct, Play) designed to support typical users aiming to accomplish certain tasks, b) the CR-PLAY mixed pipeline as an alternative approach to produce faster and cheaper high quality photorealistic content for videogames, and c) the effect of videogames, created through CR-PLAY, on the gaming experience of typical videogame players.

Three evaluation studies (aligned to the CR-PLAY software development roadmap) will be conducted which will follow a methodological approach which entails complementary evaluation methods (analytical, expert based, user and explorative). Analytical and expert-based evaluation methods will be mainly used in low fidelity evaluation studies whereas empirical and comparative evaluation studies will be mostly conducted with the high-fidelity and the final release of CR-PLAY. This is mainly because the quality of photorealistic content that is produced by CR-PLAY depends heavily on the implemented IBR/VBR algorithms and therefore this aspect can be evaluated after the high-fidelity prototypes are delivered.

A within-subject study design will be adopted in the context of the comparative evaluation studies. In such a setting the experimental group (using the CR-PLAY approach) and the control group (using the traditional approach) are made of the same people. Given the fact that the participants will need some time to learn and develop skills in using the CR-PLAY approach we will focus during the first comparative evaluation study rather on qualitative aspects of users' perspective. Therefore in this study will not focus on quantitative measures of observed performance but will rely on end-users feedback and perception with regards to improvements that are required for the CR-PLAY approach. Quantitative aspects will be examined during the final comparative evaluation study (summative study). It is anticipated that the participants in the final study will have acquired the desired skills and experience in using the CR-PLAY tools. In addition, there will be an effort to increase ecological and internal validity of the comparative evaluation studies, since the game development tasks will be integrated in a real videogame developer context and the participants will interact with both approaches at their own work environments without the intervention of any experimental equipment or the evaluators during the summative study.

Furthermore, it is important to stress that an important aspect that affects the usefulness of the CR-PLAY approach is related with the task complexity of the content creation tasks within a given game design. This issue of controlling the task difficulty, as we study comparatively the traditional with the CR-PLAY approach, is an important aspect that will be taken in consideration in the frame of the performed studies. Therefore, a cornerstone aspect of the comparative evaluations is to define representative game designs for **videogame prototypes** that will embrace content creation tasks of diverse difficulty.

Finally, typical videogame players will be also participating in the frame of comparative evaluation studies, especially in the final phase, aiming at gathering their gaming experience while interacting with videogames produced by the traditional and the CR-PLAY approach.

Table of Contents

Abbreviations and Acronyms

- **DoW:** Description of Work
- **EC:** European Commission
- **FP7:** Seventh Framework Programme
- **R&D:** Research and Development
- **UCD:** User Centred Design
- **CIF:** Common Industry format
- **KLM:** Keystroke Level Model
- **TA:** Task Analysis
- **GOMS:** Goals, Operators, Methods, and Selection rules
- **UI:** User Interfaces
- **CIT:** Critical Incidence Evaluation Technique
- **SUMI:** Software Usability Measurement Inventory
- **IMI:** Intrinsic Motivation Inventory
- **PANAS-X:** Positive Affect Negative Affect Schedule-Expanded

1. Introduction

The evaluation plan of a project involving software development is a crucial part of the user-centered design process [ISO 9241-210]. It entails a detailed roadmap with regards to evaluation studies that will be performed with typical end-users and specifies factors/criteria that will be evaluated. These evaluation criteria are measured in formative (i.e. with the aim to inform design) and summative (i.e. with the aim to assess achieved result) evaluation cycles with the aim to identify flaws and recommend improvement of the product, finally establishing a measure of success of the project. In this realm, the evaluation studies can lead the iterative design process and validate whether end-user requirements and expectations are satisfied.

According to the user centred design (UCD) approach, adopted in CR-PLAY, the evaluation of the software prototypes should start early in the development process. In our case, the evaluation plan comprises two formative evaluation studies (*during the software development process*) and one summative (*after the final release of the CR-PLAY pipeline*) evaluation study. Formative evaluation is typically conducted several times during the development process of a product. During the process often subjective measures and qualitative data are produced [1]. The aim of the formative evaluation is to find out and understand how a product or system is used by typical users with respect to expected usage and to propose recommendations for improvements. On the other hand summative evaluations, which are performed usually at the end of the product development process, focus rather in providing empirical evidence about the usefulness and usability of a software system. Summative evaluation is typically quantitative and related to performance data which might lead to general conclusions about the usefulness of a product or system.

In this context, the objective of this deliverable is to define a multilayer usability evaluation plan embedded in the iterative design and implementation of CR-PLAY pipeline. The proposed evaluation plan will consist of multiple layers describing the user interaction in different perspectives and by using different levels of abstraction taking into consideration factors/criteria that affect the user experience. The evaluation plan specifies evaluation factors/criteria, appropriate evaluation methods and techniques, and associated qualitative and quantitative measures. The applied evaluation techniques are based on user profiles and models which describe what constitutes user's cognitive, social and contextual factors in which interaction with CR-PLAY process and tools takes place. The expected results include suggestion of ways of improving the user experience and the fitness of purpose of CR-PLAY's pipeline through organizing and presenting information and functionalities in an appropriate format to diverse user groups.

The deliverable is structured as follows: in section 2 we provide an overview of the most relevant evaluation methodologies. This section will serve as a point of reference that explains in detail the methodologies that will be used during the project lifecycle. Section 3 describes the main evaluation factors that are considered important in the frame of the CR-PLAY project. The evaluation factors (from an end-users point of view) were derived from the CR-PLAY description of work and the deliverable "D4.1: End-user requirements analysis". In sections 4 and 5 we provide some further insight in the planned evaluation studies, bearing in mind that further refinement of the procedures and users involved will be defined further on during the process as these evaluation studies need to be aligned to CR-PLAY's software development evolving roadmap. Finally, in section 6 we provide examples of the evaluation reports that will be produced. The produced evaluation reports will follow the conventions and structure proposed in the Common Industry Format (CIF). CIF is a standardized way to produce usability evaluation reports based on the ANSI-NCITS 354-2001and ISO/IEC TR 25062 standard. The findings of the evaluation studies along

with recommendations for improvements will drive the development of the CR-PLAY mixed pipeline for videogames development.

2. Background

A variety of methods and techniques that support the evaluation of software systems and products have been proposed and used. A detailed review of current state of the art evaluation methods and techniques is out of the scope of this deliverable. However we concisely describe the main evaluation methodologies and associated techniques that can be applied within the CR-PLAY project. These methodologies are categorized in four broad categories: (a) usability inspection, (b) analytic methods (c) user testing, and (d) exploratory methods.

2.1 Inspection methods

Inspection methods are used by evaluation professionals (the “experts”) to inspect, evaluate, and assess the usability of a system, without involving typical users. Although the basic usability criterion is the effective use by typical users, this principle does not reduce the usefulness of evaluation methods based on experts’ evaluation approaches. According to these methods, evaluators who have expertise on rules and methodologies for evaluation of interactive systems emulate the expected formal use of the system. These methods can have a formative character, because they can be applied during initial stages of an iterative design cycle, with comparatively lower cost than the observation of users.

Numerous inspection evaluation methods have been proposed, including heuristic evaluation [2], cognitive walkthrough [3], usability walkthrough [4] and applications of guidelines in walkthroughs [5] etc. Among these, the most well-known and frequently used methods are heuristic evaluation and cognitive walkthrough. These two evaluation methods will be used mainly in the formative evaluation studies (low- and high-fidelity) of CR-PLAY.

Heuristic evaluation [2], involves usability specialists who judge whether an interactive system or product follows established usability principles the "heuristics" (typical heuristics are included in Annex 2). Heuristic evaluation pays special attention to ensure proper design of interactive products in terms of layout consistency, appropriate feedback and minimizing errors. Heuristic evaluation is usually performed by having each individual evaluator inspect the product alone and make suggestions for improving the design. Only after all evaluations have been completed the evaluators are allowed to communicate and have their findings aggregated. This procedure is important in order to ensure independent and unbiased evaluations from each evaluator. Each evaluator delivers a written report. Written reports need to be read and aggregated by an evaluation manager.

Cognitive walkthrough simulates in detail the process of carrying out specific tasks, in the flow of interaction with the system, determining if the simulated objectives of user and the system’s response, can theoretically lead to the next required operations. The usability evaluator goes through the system simulating a typical user cognitive state and performs tasks that the user typically performs. The method focuses on ease of learning. It is designed to give designers the ability to identify problems associated with learnability before an implementation of the design, prior to having users test it. In a cognitive walkthrough, evaluators "walk through" the steps users need to take to accomplish a goal. A cognitive walkthrough requires the preparation of a set of tasks that are representative of those that users will perform using the interactive tools, along with the sequence of actions to accomplish the tasks and a description of the interface (a functional specification, a paper mock-up, a paper prototype, an interactive prototype, or a prototype working system).

2.2 Analytical methods

Analytical methods concern the implementation of models that represent the behavior of users with a given technology. These models describe goals and sequences of activities that typical users executes during their interaction with the technology. The models in these methods are distinguished in predictive models that are produced by the designer (*designers' models*) and descriptive models that are produced after observation of actual system usage (*user's models*). A typical method of the first category (predictive user models) is Keystroke Level Model (KLM) [8] and of the latter category (descriptive user models) is the Task Analysis (TA) method [9].

By using this approach our intention is to build a conceptual model, referring to the way the users view the system and perform tasks for accomplishing certain goals. The objective is to produce a clear understanding and detailed analysis of how a task is accomplished, including a detailed description of both manual and mental activities, task durations, task frequency, task allocation, task complexity, environmental conditions, necessary equipment, and any other unique factors involved in or required for one or more actors to perform a given task. Having these user interaction models ready we will be able subsequently to reflect on the observable behaviour of users captured by user testing and to identify bottlenecks and design interaction flows in the human-machine interaction.

2.3 User testing methods

User testing methods involve testing a product or a system with targeted users performing intended tasks. User testing is the only method which provides empirical data (objective) with regards to user interaction. The goal is to observe and analyze users' understandings and misunderstandings, successes and failures, correct actions and errors, and task completion times to infer potential user interaction problems of the product or system. User testing methods often embrace a simple research design, and a small number of subjects, with a quick turnaround time. User testing studies are usually conducted in a controlled area that is specially set up for that purpose, such as a usability laboratory with video, audio recording equipments, eye trackers for monitoring attentional resources and one-way mirrors across the observers' area. However, in cases in which either typical end-users are difficult to be moved to a usability laboratory or the nature of the tasks does not allow for such an approach, the use of remote usability evaluation techniques can be a viable alternative [6].

In the context of user studies, a well-known evaluation technique is related to the thinking aloud protocol [7]. During this the participant is asked to vocalize her/his thoughts, feelings, and opinions while interacting with the software, performing a typical part of a task of a user scenario. The main advantages of the method is its feasibility to collect qualitative feedback directly from the users, the support for comprehension of the conceptual model of the user, concerning the interaction process and the terminology the user uses in order to express an activity, which should be incorporated in the design. In the frame of the CR-PLAY we will apply, in a remote context, the think aloud protocol during the formative evaluation studies in order to assess end-users views with regards to the proposed CR-PLAY tools. Furthermore, a comparative eye-tracking study with typical game players is planned to be conducted by UPAT in order to track their attention, related to emotional state during their interactions with videogames prototypes created through the traditional and the CR-PLAY approach.

2.4 Exploratory methods

Exploratory evaluation methods aim at providing feedback about the preferences, the needs and the particularities of the users, by giving them the occasion to express their opinion. These methods include techniques as: a) Critical Incident Methods, Contextual Inquiry, Diary studies b) Focus Groups, c) Questionnaires and d) Interviews. In the frame of the performed evaluation studies we will use various

approaches in order to elicit the users views with regards to the tools used. Semi-structured interviews or focus groups will be used after the presentation of the low- and high-fidelity prototypes to representative users. Furthermore, questionnaires will be used in order to assess the user satisfaction and players immersion.

3. CR-PLAY: Evaluation plan

3.1 Methodology

The CR-PLAY project adopts a *User Centred Design* (*UCD*) approach [ISO 9241-210] which includes the specification of end-user requirements, the design and development of software prototypes and the implementation of evaluation activities (Figure 1). The objective of the evaluation activities is to examine whether the user requirements have been sufficiently met or further enhancements of the software tools and proposed processes are required. The evaluation activities in CR-PLAY rely on an evaluation plan which identify and prioritize several evaluation factors, evaluation goals and metrics.

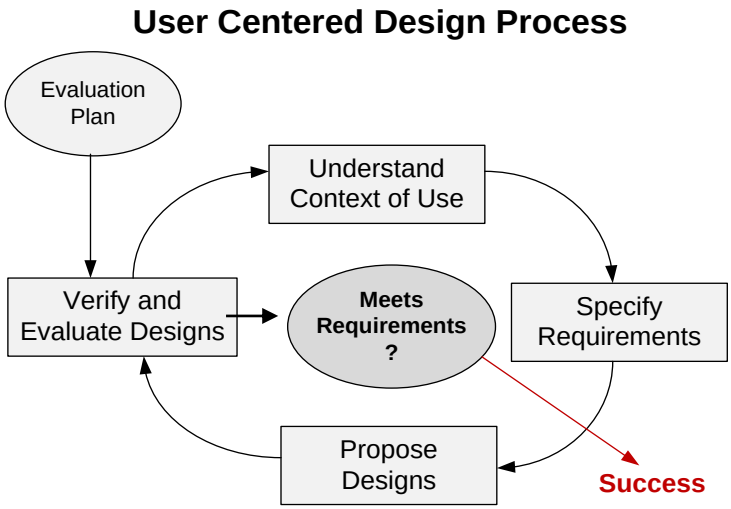

Figure 1. User-centered design process

In this context, the methodology of defining and implementing the CR-PLAY evaluation plan consists of the following steps:

1. **Specify evaluation dimensions and goals for each representative CR-PLAY user category**

   Identify, analyze and prioritize factors (dimensions) affecting user experience related to the CR-PLAY pipeline. In addition, identify evaluation objectives (goals) related to CR-PLAY by taking into consideration representative user categories and scenarios of use.

2. **Define qualitative and quantitative metrics**

   The objective of this step is to define qualitative and quantitative metrics for each identified CR-PLAY evaluation goal.

3. **Select appropriate evaluation methods and techniques**

   Appropriate evaluation methods and techniques will be selected that will specify how the evaluation studies will be conducted. These evaluation methods will be selected by taking into consideration various CR-PLAY software releases (low-, high-fidelity prototypes or final release).

4. **Implement the evaluation activities and report evaluation results**

Finally, evaluation activities will be conducted, analyzed and reported (along with suggestions for improvements) aiming to drive the iterative development process adopted in CR-PLAY.

3.2 Typical user categories and usage scenarios

The user categories of CR-PLAY along with typical usage scenarios have been identified and analyzed in deliverable "D4.1 End-user requirements analysis" (*section 3*: Stakeholder analysis, *section 9.4*: Personas, *section 7.4*: Usage Scenarios).

As depicted in Figure 2, the technical specialists are the primary users (mainly artists and programmers) and the CR-PLAY tools are built foremost for them. The primary users are mainly the ones who will directly interact with the CR-PLAY tools (i.e. capturing, reconstruction and editing tools). However, there are also secondary users (the art directors, game designers or game producers, publishers) who evaluate the produced outcome and provide initial and iterative directions with regards to the real world scenes that need to be captured and perform an overall quality control of the produced results. Finally, typical game players are also considered as CR-PLAY secondary users as they are concerned with the CR-PLAY results and are going to evaluate the quality of the produced assets. Assessing the game player's opinions with regards to the quality of the content that is produced with CR-PLAY and investigating whether their overall player (gaming) experience is affected by this new approach is considered to be an important evaluation factor.

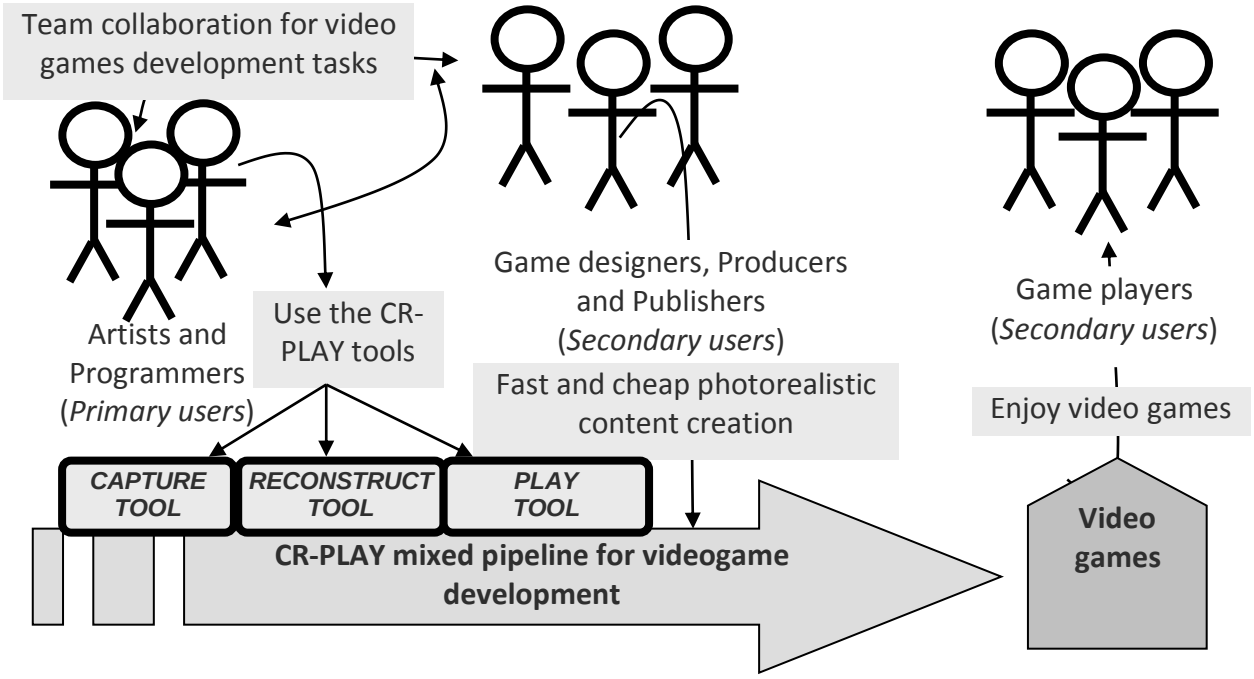

Figure 2. CR-PLAY User categories

3.3 Evaluation dimensions and metrics

The CR-PLAY mixed pipeline for videogame development aims to support a multidisciplinary activity, which embraces complex communication and collaboration tasks among several team members who share different roles within the videogame development pipeline. In this realm, it is necessary that the proposed CR-PLAY approach will be evaluated from several perspectives and diverse users' points of view. These evaluation factors are:

a) The **CR-PLAY tools** themselves as separate tools designed to support typical users aiming to accomplish certain tasks and achieve certain goals (*mainly from the perspective of the technical specialist group*).

b) The **CR-PLAY mixed pipeline** as a collaboration process that involves diverse users with different roles in the content creation pipeline of videogames (*mainly from the perspective of the technical specialist group and the game designers*).

c) The objective through the CR-PLAY pipeline to produce **faster and cheaper high quality and more creative videogames** (*mainly from the perspective of the game designers, producers*).

d) The quality output of the produced CR-PLAY videogames and its effect on **players gaming experience** (*mainly from the perspective of the game players*).

Figure 3 depicts the evaluation factors along with defined evaluation goals and specific metrics. The evaluation goals with regards to the CR-PLAY pipeline and the tools (*e.g. support for team collaboration, capturing and reconstruction performance, quality control of reconstruction, interoperability, modifiability of assets, quality of output, quality control, learnability etc.*) have been derived from the end-user requirements analysis deliverable. These evaluation goals will be set in the case of evaluation studies, which are described in more detail in the next sections.

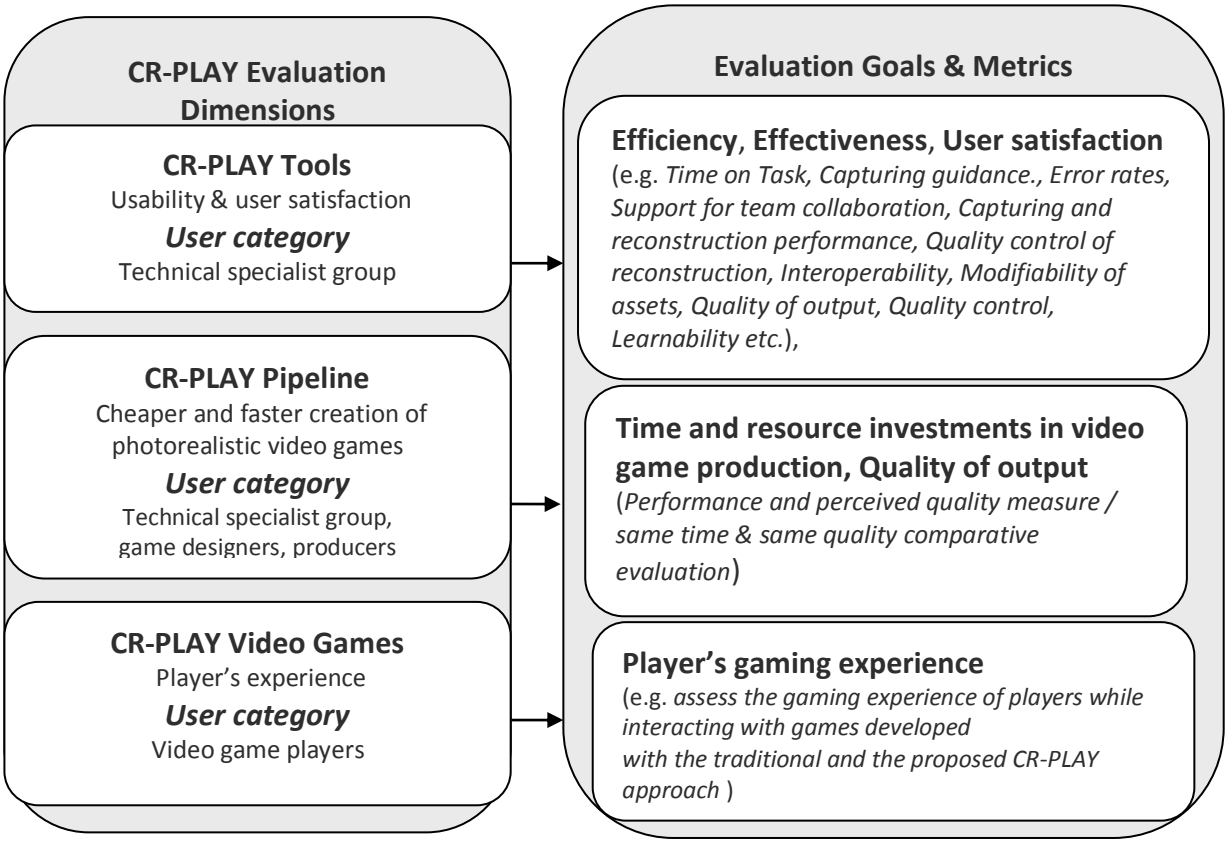

Figure 3. CR-PLAY Evaluation dimensions, goals and metrics

4. CR-PLAY: Formative evaluation studies

4.1 Scope

The objective of the formative evaluation is to provide feedback with regards to the CR-PLAY low- and high-fidelity prototypes. These prototypes will try to convey the conceptual design and will show how the main user scenarios will be supported by the CR-PLAY tools, without the necessity to have a full implementation of all functionalities. The feedback that will be derived from the formative evaluation studies will be related to these conceptual decisions with regards to the user interface, interaction design, information architecture, supporting material and the supported functionalities.

4.2 Method

The formative evaluation studies will follow a methodological approach which entails complementary evaluation methods (analytical, expert, user and explorative). Analytical and expert evaluation methods will be used mainly in low-fidelity evaluation tasks whereas user empirical studies will be mostly conducted with high-fidelity prototypes. Field studies (i.e. contextual inquiry) will be conducted during the summative evaluation phase and are not scheduled within the formative evaluation phase of CR-PLAY.

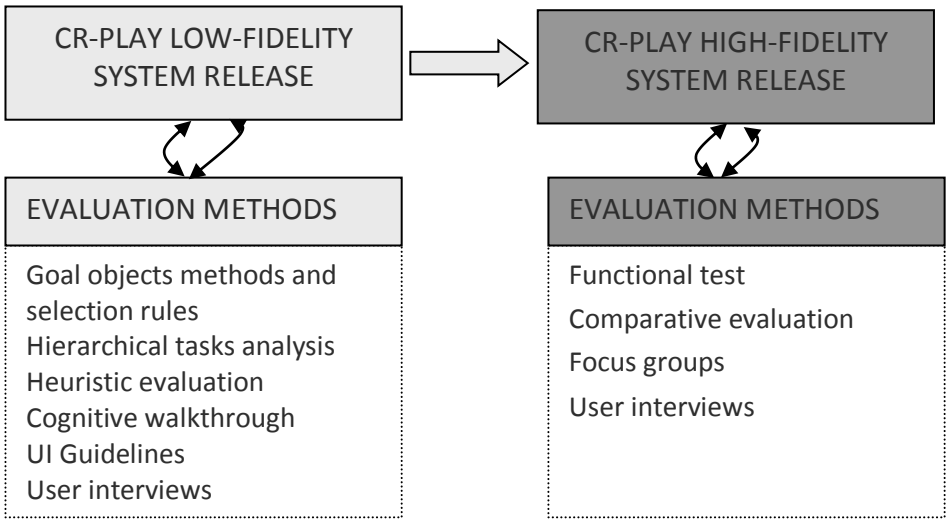

Figure 4. CR-PLAY Evaluation methods applied in formative studies

Low-fidelity evaluation

Depending on the maturity of the delivered low-fidelity prototypes, user tasks will be identified and analytical evaluation methods (Hierarchical Task Analysis, GOMS, and Action Analysis) will be performed. By using this approach our intention is to build a conceptual model, referring to the way the users are expected to view the system and perform tasks for accomplishing certain goals. The objective is to produce a clear understanding and detailed analysis of how a task is accomplished, including a detailed description of both manual and mental activities, task durations, task frequency, task allocation, task complexity, environmental conditions, necessary equipment, and any other unique factors involved in or required for one or more actors to perform a given task. Having these user interaction models ready we will be able at a later time to reflect on the observable behavior of users captured by user testing and to identify bottlenecks and interaction flows in the human-tools interaction.

The next step of the low-fidelity evaluation will involve expert based evaluation methods (heuristic evaluation, cognitive walkthrough, and application of UI guidelines). The heuristic evaluation will be based

on heuristics presented in Annex 2, while UI guidelines are more specific to the CR-PLAY technology. The main goal of expert evaluation can be summarized as follows: a) Identify and reduce the number and severity of identified usability problems, b) Use results as a reference point to plan end-user empirical studies of the next phase and c) initial assessment of user experience aspect of prototypes and functional services. During this evaluation phase, the experts will meet in the frame of group-based expert evaluation panels in order to discuss findings, agree on their severity and propose recommendations for increasing the overall user experience of the CR-PLAY mixed pipeline.

Furthermore, typical end-users (technical specialists) will be involved in reviewing the low-fidelity CR-PLAY software prototypes. In the case of the formative evaluation studies the empirical studies will be designed to gather rather qualitative than quantitative feedback; that is to provide useful insights on the interaction design of the CR-PLAY tools. The evaluation protocol that will be defined by the experts will specify the user roles, tools and typical user tasks. Remote evaluation [6] will be considered as a viable alternative to communicate with typical users while interacting with the delivered prototypes. A think-aloud protocol will be applied in the frame of the latter studies. Furthermore, focus group studies will be applied aiming to elicit user subjective feedback and experience that had while interacting with the delivered prototypes.

High-fidelity evaluation

In this phase of the evaluation process, which is planned after the high-fidelity prototypes will be delivered, the evaluation will involve first expert evaluation, like in the previous phase, and in addition a number of user studies. The latter will include a comparative study between the traditional and the CR-PLAY mixed pipeline for videogames development. Comparative evaluations will be applied on two evaluation criteria: a) performance and b) quality. As stated in the DoW (p.21): "*Same time evaluation refers to efficiency and effectiveness when game developers create assets with traditional methods and with CR-PLAY tools on specific tasks, computational resources and time frames whereas same quality evaluation refers to the gamers experience relating to the quality of the produced outcome.*"

The evaluation studies with the high-fidelity prototypes of CR-PLAY will drive the final implementation of the mixed CR-PLAY pipeline and therefore they must provide feedback with respect to a series of important issues. These issues will be related to:

- How easy it is for a game developer to adapt the traditional working flow (methodology and tools) towards integrating the proposed CR-PLAY method and tools.
- How easy it is for a game developer to understand the proposed approach and use the CR-PLAY tools. How fast can typical users learn to use the CR-PLAY tools in order to accomplish certain tasks. How well are the typical tasks (capture, reconstruct, play) supported by the CR-PLAY tools?
- How well the terminology used in the CR-PLAY tools match the game developer's vocabulary and/or meet their expectations?
- How easily the outputs of the CR-PLAY tools are integrated in state of the art game engines and/or modelling tools, already used by typical users. Do the CR-PLAY objects behave in these environments as expected by the end-users?
- Are the CR-PLAY tools more cost effective compared to traditional tools for game development? At the same time how the quality of the content produced with the CR-PLAY tools is compared to the ones created with traditional techniques?

- What is the typical players' experience when playing games created with CR-PLAY? Does the CR-PLAY approach support the creation of more creative videogames (e.g. are videogames embracing known graphical content like streets or buildings more immersive than videogames that contain content the users are not familiar with, can new genres of games be produced with the new assets? Can existing games change with the possibilities of introducing familiar assets in the game world?)?

Evaluation design

Aiming to address the aforementioned evaluation goals, a detailed evaluation plan, tailored to the available resources needs to be defined. This is a prerequisite in order to assure validity of results and minimizing of biases like the learning effects of the CR-PLAY tools on the users. The first aspect of a good evaluation design is related to the ecological validity of the plan, which requires that the experimental design, procedure and setting of the study must be shaped by the real-life context of the typical users' environment. A second aspect is related to the internal validity of the study, which means that typical users will interact with the two approaches in order to develop typical videogames prototypes of varying complexity, under the same conditions, available time frames and computational resources. Finally, another important factor is related with the external validity, which determines whether the conclusions drawn can be generalized to a wider extent beyond the specific study.

In this context, aiming to increase internal validity we will select participants with long experience in the content creation and videogames development. There will be also an effort to increase ecological validity of the study since the game development tasks will be integrated in an authentic videogame developer context and the participants will interact with both approaches at their own physical environments without the intervention of any experimental equipment or other intermediaries. During the first comparative evaluation study the focus will be on gathering end-user experiences in using the CR-PLAY approach and tools. Quantitative data are not applicable during this study since the participants involved will be experienced users with respect to traditional pipeline but novice users with the new CR-PLAY mixed pipeline, while the experimental nature of the evaluated prototypes will not allow for measuring performance.

In this context, an important decision with regards to the comparative evaluation studies is whether a within-subjects or between-subjects experimental design should be adopted. In a between subject study design the experimental user group (users trying the CR-PLAY approach) and the control group (following the traditional approach) consists of different people. On the other hand, in a within-subject design the experimental and the control group consist of the same people who try both approaches in arbitrary order. In the context of CR-PLAY, a within-subject design fits better to the comparative evaluation goals since the research is focused rather on the two alternative approaches and not on individual differences that might affect the assets creation tasks (e.g. participants with greater experience might be more efficient than others).

However, in the context of a within-subjects design another important issue is whether the experimental and control group can proceed with the same videogame prototype creation or whether two different videogames prototypes, sharing the same task difficulty with regards to content creation, are required to be developed. The concern is that the participants might maintain a bias derived from their previous experience in creating the same content and this might have an effect on efficiency, effectiveness, overall engagement etc.

On the other hand, providing the study participants with two different content creation tasks has some limitations. Firstly, it is very difficult to define two different content creation tasks with the same difficulty level. In addition, it will be almost impossible to compare the produced quality of the assets if the game design task will not be the same.

Therefore, based on the evaluation goals of CR-PLAY a valid within subject's experimental design requires that the participants will accomplish the creation **of one videogame prototype** with the traditional and the CR-PLAY approach. In particular half of the developers will accomplish the creation of the videogame prototype with CR-PLAY and will then switch the method and do the prototype again with the traditional method. In this context, aiming to minimize the previous experience effect with regards to the method used we will randomly ask half of the game developers to start with the CR-PLAY method and the other half with the traditional method.

Pre-Study:

- Step (a): Produce a game design document that will drive the development of one videogame prototype in the frame of the comparative evaluation studies. In context, the game design document will include several tasks of content creation. It should involve easy, average and difficult content creation tasks for a given videogame. In a game design in which the content creation task is too easy it will probably lead that both approaches (CR-PLAY and traditional) will have no variations with regards to the evaluation goals.

- Step (b): For each videogame developer, choose whether he will start with the CR-PLAY or the traditional approach. Thus, we aim to eliminate the previous-experience factor.

Study Implementation:

- Step (a): Given that the videogame development is a collaborative activity that requires considerable amount of time, it is very difficult to observe and evaluate this process in real time. Therefore, during the implementation of the study the videogame development team will implement a critical incidence evaluation technique (CIT). This technique will allow us to conduct the evaluation studies without direct observation of the users. Critical incidents are related to positive or negative user experiences that affect task performance. The participants will report, using a given template, incidents about what actually happened (before, during and after the incident).

Post-Study:

- Step (a): Semi-structured interviews will be performed with the participants. The emphasis will be on understanding in detail the positive or negative user experiences in using the high-fidelity prototypes in terms of efficiency, effectiveness and quality of output.

- Step (b): Focus group study will be organized aiming to discuss the hands on experience of the users in using the CR-PLAY tools and provide feedback that can be used for improvements with regards to the final release of the CR-PLAY mixed pipeline. The suggestions for improvements will focus on efficiency, effectiveness and quality of output.

In the frame of the formative evaluation studies, the aforementioned research questions will be answered based on the described evaluation design and they will be of qualitative nature and related suggestions for improving the CR-PLAY process and tools. These studies will not gather yet quantitative measures (since this will not be the final CR-PLAY version) but will rely on end-users qualitative feedback and views

concerning improvements that need to be done in the final version. It is anticipated that these improvements will affect the speed and quality of output of the final CR-PLAY release.

4.3 Participants

The analytical- and expert-based evaluations will be conducted by three experts who will be recruited by UPAT. With regards to the empirical evaluation studies three (3) to five (5) representative game developers will be recruited by the CR-PLAY consortium. The recruitment is expected to be done from the game developers that participated in the end-user requirements analysis (Deliverable 4.1). In particular, Testaluna (TL) and game companies associated with CUR will be the main participants in the frame of the evaluation studies, in similar fashion as in the Requirements Gathering phase of the project. Furthermore, other game developers will be contacted with the aim to be able to perform timely parts of the evaluation studies. Given that the high-fidelity comparative evaluation study will demand of the game developers considerable time and human resources, the details of the study cannot be decided yet. Therefore, the game developers and the precise definition of game design tasks are considered to be crucial aspects that affects the implementation of the comparative evaluation study and are beyond the scope of the current document to specify in more detail at this early stage of the project. It is expected that the consortium management committee will decide on the resources and the exact nature of the tasks of these studies in appropriate time.

5. CR-PLAY: Summative evaluation studies

5.1 Scope

The objective of the summative evaluation study is to perform further comparative evaluations between the traditional and the final mixed CR-PLAY pipeline for videogames development, in order to measure the effectiveness and value of the proposed technologies. These evaluation studies will focus on both quantitative and qualitative measures aiming to provide empirical evidence that the CR-PLAY pipeline can deliver faster and cheaper videogames with acceptable quality photorealistic content. Typical game players will also participate in the frame of the CR-PLAY summative evaluation studies aiming to investigate the acceptance of the CR-PLAY produced videogames and the CR-PLAY's effect on the gaming experiences.

5.2 Method

The objective of the summative evaluation studies is mainly to provide evidence that the CR-PLAY mixed pipeline of videogame development is a viable alternative to current industry practices with a competitive advantage which relies on delivering faster and cheaper high quality content. Therefore this part of the evaluation will be solely based on user studies. These comparative evaluation studies will be performed by following a within-subjects design as it has been described in the formative evaluation studies.

However, in the summative evaluation phase of CR-PLAY, another important factor that will be examined and can influence the final conclusions with respect to the usefulness of the CR-PLAY approach is related with the varying degree of task difficulty of the content creation tasks of videogames. This issue of controlling the game design difficulty, as we study comparatively the traditional and the CR-PLAY approach, is an important aspect that needs to be evaluated. The aim is to evaluate the speed, required budget and produced quality of output in several game designs in which the content that needs to be captured, reconstructed and integrated into a videogame will vary. From a game design perspective, the given task for content creation should be difficult enough to challenge experts but at the same time will be applicable as well for novice users.

Several data about the user interactions will be collected throughout the study.

Performance measures: Performance measures will be related to task efficiency and effectiveness and will provide objective data on whether the content creation tasks with the CR-PLAY method are faster comparing to the traditional method. Some examples of these measures are the following: *total time required for successful task accomplishment, total attempts required for successful task accomplishment, the number of cases in which the participants were not able to complete a task due to an error from which they could not recover, the average amount of time to complete each task etc.*

User Satisfaction: User satisfaction questionnaires will be deployed such as the SUMI - Software Usability Measurement Inventory, a 50 item questionnaire that measures five aspects of user satisfaction (Likability, Efficiency, Helpfulness, Control and Learnability), scoring the products against expected industry norms.

User Preference & Quality of output: Qualitative data will be collected at the end of the study through semi-structured interviews and focus groups, to elicit the users' subjective preference and perceptions regarding the interactions they had with both approaches and the perceived quality of the content.

Evaluation with typical game players

This phase of the CR-PLAY evaluation will involve typical players aiming to gather their gaming experience, which is derived from their interaction with videogames prototypes produced with both approaches. As stated in the DoW (p.21): "*With regards to evaluating the quality of output we will conduct user studies that will involve gamers in order to assess their gaming experience while interacting with games developed with the traditional and the proposed approaches.*"

In this context, investigating the relationship between the gaming experience and the quality of the content which is produced by traditional and CR-PLAY method is a complex endeavour. A plethora of constructs exist that measure the effect of different factors on gaming experience (*e.g. immersion, fun, presence, involvement, engagement, flow, content quality, playability, graphics, sound, narrative, story, rules and controls as imprinted elements etc.*). Accordingly, several frameworks have been proposed for evaluating the gaming experience and instruments (mainly questionnaires) have been developed to assess the players gaming experience for a given videogame. A recent game experience questionnaire proposed by [10] proposes to measure the game experience using several factors (*e.g. enjoyment, frustration, puppetry, game control, facilitators, ownership, game-play and environment*).

Other studies investigate how graphic quality affects players gaming experience. Bracken et al. [11] investigated the effects of image quality in immersion and presence and reports a significant effect on image quality to player's immersion. A more recent study [12] revealed that high-fidelity graphics in casual games result in a more positive impression of the game. To evaluate player's experience they reported on the interest-enjoyment subscale of the Intrinsic Motivation Inventory (IMI) which reflects intrinsic motivation and the PANAS-X (Positive Affect Negative Affect Schedule-Expanded) questionnaire. This study concludes with two interesting results. One is that graphical fidelity has an impact on player experience when graphics are integrated with game mechanics and players attend to them during game play and the second is that that the way graphical fidelity affects player experience is influenced by difficulty. The most challenging mechanics a game includes the better graphical fidelity it needs to offer in order to achieve the same level of player experience that easier games provide with low-fidelity graphics.

In the frame of the CR-PLAY project we will investigate the gaming experience of typical game players through a subjective (self-reporting) and objective approach (user testing). These two approaches are complementary in nature and entail several advantages and disadvantages.

A comparative eye-tracking study with typical games players will be conducted by UPAT in order to track their attentional resources while interacting with two variations (with regards to the content creation) of the same videogame, associating their behaviour to engagement and other constructs relating to player satisfaction. UPAT runs a fully-equipped usability laboratory, including an eye tracking device together with in-house developed software for data analysis. This has already been used in a number of quantitative analysis of user behaviour in tasks like information search of varying degree of difficulty [13].

The within-subjects experimental design will initially ask typical game players to play (for about 20 minutes) two versions of the same videogame prototype. The two game versions will differentiate with regards to the content creation approach and half of the participants will initially play with the version created through the traditional approach and afterwards they will play with the CR-PLAY version of the game. The rest of the participants will play the other way around. Through the proposed eye-tracking study we will investigate on which parts of the game the players are looking, what part of the content triggers their attention and for how long, how their focus moves within the game-play etc. The produced results from the eye tracking analysis (*like heat maps, saccade pathways, number of fixations*, *number of fixations per second etc.*) will be comparatively evaluated and conclusions will be driven with regards to whether and how the quality of the content affects the attention of the players. Finally, the immersion of the players will be acquired, on a subjective basis, through the use of semi-structured interviews and/or questionnaires.

5.3 Participants

With regards to the empirical evaluation studies, three to five representative game developers will be selected by the CR-PLAY consortium. It is important to stress that the game developers that will participate in this phase of the evaluation should have acquired the necessary experience in using the CR-PLAY tools prior of the study. In particular, Testaluna (TL) and game companies associated with CUR will be the main participants in the frame of the summative evaluation studies. Furthermore, it is important to mention that the participants of TL should not be the same people who will design and develop the CR-PLAY tools on behalf of TL. Finally, for the game players' evaluation study we will recruit a number between 10 to 20 game players, mainly gamers from the University of Patras.

6. Structure of evaluation reports

The Usability reports that will be produced during the proposed evaluation plan will follow the conventions and structure proposed in the Common Industry Format (CIF). CIF is a standardized way to produce usability evaluation reports. It is an ANSI-NCITS 354-2001and ISO/IEC TR 25062 standard. This standard is mainly used in cases of summative studies that involve representative users interacting with the evaluated system, in which typical quantitative measures of task success and number of errors are counted. However, it is often used in formative studies, as it provides a standardized way to describe the methodologies employed and report usability findings.

A report that is produced according to the aforementioned standard should include the following basic sections: 1) Summary, 2) Introduction, 3) Methods and 4) Results. It is important to note that this standard assumes that the best practices of user-centred design interaction are employed, as described in the standards ISO 13407 and ISO 9241-11.

The identified usability flaws will be classified according to their importance in the following four categories:

- ***Catastrophe:*** The identified issue is so severe, that the user will not be able to complete the task, and may not want to continue using a CR-PLAY tool.

- ***Major:*** Users can accomplish the task but only with considerable frustration and performance of unnecessary steps. The user will have great difficulty in circumventing the problem.
- ***Moderate:*** The user will be able to complete the task in most cases, but will have to undertake some moderate effort in getting around the problem. They may need to investigate several links or pathways through the system to determine which option will allow them to accomplish the intended task. Users will most likely remember how to perform the task on subsequent encounters with the system.
- ***Minor:*** The problem occurs only intermittently, can be circumvented easily but can be irritating. It could also be a cosmetic flaw.

7. Conclusion

The current deliverable presents the evaluation plan of CR-PLAY. The proposed evaluation will take place in three phases, both in a formative and summative form and is based on a variety of quantitative and qualitative methods and techniques. In general, no single evaluation method can unveil all possible user interaction flaws and depending on the maturity level of the delivered prototypes specific evaluation techniques need to be applied. As it has been mentioned in detail in the deliverable, analytical and expert evaluation methods will be mainly used in low-fidelity evaluation studies whereas empirical and comparative evaluation studies involving typical users will be mostly conducted with the high-fidelity prototype and the final release of CR-PLAY tools. The main objective of the formative evaluation studies will be to gather qualitative feedback from typical end-users that will be used as an input for the subsequent CR-PLAY releases. On the other hand, the objective of the summative evaluation studies will be to gather qualitative as well as quantitative information with the aim to determine whether the CR-PLAY pipeline provides a viable alternative for creating high quality, faster and cost effective videogame assets as stated in the objectives of the project.

References

[1] Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.

[2] Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen & R. L. Mack (Eds.), *Usability inspection methods* (pp. 25-64). New York: Wiley.

[3] Lewis, C., & Wharton, C. (1997). Cognitive walkthroughs. In M. Helander, T. K. Landauer, & R Prabhu (Eds.), *Handbook of human-computer interaction* (2nd ed., pp. 717-732). Amsterdam: Elsevier Science.

[4] KaratJ., &Bennett,J. L. (1990). Supporting effective and efficient design meetings. Proceedings of IFIP *INTERACT'90 Third International Conference on Human-Computer Interaction*, 365-370.

[5] Jeffries, R., Miller, J. R., Wharton, C., & Uyeda, K. M. (1991). User interface evaluation in the real world: A comparison of four techniques. *Proceedings of ACM CHI 91 Conference on Human Factors in Computing Systems,* 119-124.

[6] C.A. Fidas, C. Katsanos, E. Papachristos, N. Tselios, N. Avouris (2007). "Remote Usability Evaluation Methods and Tools: A Survey". In T. Papatheodorou, D. Christodoulakis and N. Karanikolas (Eds.) Proceedings of PCI 2007 (pp. 151-163), Patras, Greece, 18-20 May, 2007, ISBN: 978-960-89784-0-9.

[7] Lewis, C. (1982). Using the 'thinking-aloud' method in cognitive interface design. (IBM Research Report RC 9265, 2/17/82). Yorktown Heights, NY: IBM T.J. Watson Research Center.

[8] Card, S., Moran, T, & Newell, A. (1983). *The psychology of human computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

[9] Kieras, D. (1996). Task analysis and the design of functionality, CRC Handbook of Computer Science and Engineering, CRC Press.

[10] Calvillo Gamez, E., Cairns, P., Blandford, A., 2008, Assessing the Gaming Experience using Puppetry.

[11] Bracken, C.C., Skalski, P.,2009 Telepresence and video game the impact of image quality.

[12] Gerling, K.,Birk, M., Mandryk, R., Doucette, A., 2013, The Effects of Graphical Fidelity on Player Experience.

[13] Katsanos C., Tselios N., Avouris N. (2010). Evaluating website navigability: Validation of a tool-based approach through two eye-tracking user studies. New Review of Hypermedia and Multimedia, 16(1&2), p.p. 195-214

Glossary

For the purposes of this document, the following definitions apply.

Context of use
The users, tasks, equipment (hardware, software and materials), and the physical and social environments in which a product is used [*ISO 9241-11:1998*]
The physical environment includes the workplace design (furniture, posture and location) and the visual, auditory, thermal and atmospheric conditions, and any health and safety hazards. The social environment includes social and organizational issues including the work organization and structure, availability of assistance and interruptions, presence of other people, job design and autonomy.

Requirements
Expression of a perceived need that something be accomplished or realized
[*ISO/IEC FCD 25030*]

User
The person who interacts with the product
[*ISO 9241-11:1998*]

User group
Subset of intended users that are differentiated from other intended users by factors such as age, culture, knowledge, skill, expertise, role or responsibility that is likely to influence usability. This may include current users, potential users, users with disabilities, expected future users, users of the task output, and staff who support or maintain the product.

Scenarios of use
How users carry out their tasks in a specified context.

Stakeholder
A party having a right, share or claim in a system or in its possession of characteristics that meet that party's needs and expectations
[*ISO 15288:2002*]

Goal
An intended outcome.
[*ISO 9241-11:1998*]

Task
The activities required to achieve a goal. These activities can be physical or cognitive. Job responsibilities can determine goals and tasks.
[*ISO 9241-11:1998*]

Usability
The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.
[*ISO 9241-11:1998*]

Effectiveness
The accuracy and completeness with which users achieve specified goals
[*ISO 9241-11:1998*]

Efficiency
The resources expended in relation to the accuracy and completeness with which users achieve goals
[*ISO 9241-11:1998*]

Annex 1. CR-PLAY evaluation roadmap

Month	Deliverable / Milestone	Aim
M12	D1.1. Initial report describing capture and reconstruction Tools *MS1.Initial version of capture and reconstruction tools*	A first capture tool for a restricted setting is available. Post-processing is possible in a separate system. Data can be fed into the CR-PLAY pipeline.
M12	D4.3. Low-fidelity prototypes of mixed pipeline for videogame development. *MS10. Initial implementation (low-fidelity prototype) of mixed pipeline for videogame development*	A first integration of IBR and VBR is available. Its features and functionalities can be tested by tool developers and game developers. Formative evaluation activities can be performed.
M12 - M14	MS13. Results of low-fidelity prototype evaluation and recommendations for next design iteration	To perform the first formative evaluation study and provide feedback for the next design iteration.
M18	D1.2. Initial report describing planning and guidance system *MS2. Initial version of planning and guidance system*	A capture tool with planning and guidance is available and initial HCI evaluations have been performed.
M22	D4.4 .High-fidelity prototypes of mixed pipeline for videogame development *MS11. Advanced implementation (high-fidelity prototype) of mixed pipeline for videogame development*	An advanced integration of IBR and VBR is available. Editing and rendering functionalities can be tested by tool developers and game developers. Formative evaluation activities can be performed.
M22 - M26	D5.2. Formative evaluation of high-fidelity prototypes MS14. Results of high-fidelity prototype evaluation and recommendations for next design iteration	To perform the second formative evaluation study and provide feedback for the next design iteration.
M33	D4.5. Final implementation of mixed pipeline for videogame development and report on technical validation *MS12. Consolidated implementation (final prototype) of mixed pipeline for videogame development*	A consolidated prototype of the system is available. Its features and functionalities can be tested and used by tool developers and game developers. Summative evaluation activities can be performed.
M33 - M36	D5.3 Summative evaluation	To comparatively evaluate the traditional against the mixed pipeline of videogames development.

Table A.1. CR-PLAY evaluation roadmap

Annex 2. List of heuristics

Visibility of system status
The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.
Match between system and the real world
The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.
User control and freedom
Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.
Consistency and standards
Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.
Error prevention
Even better than good error messages is a careful design which prevents a problem from occurring in the first place.
Recognition rather than recall
Make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.
Flexibility and efficiency of use
Accelerators -- unseen by the novice user -- may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.
Aesthetic and minimalist design
Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.
Help users recognize, diagnose, and recover from errors
Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
Help and documentation
Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

Table A.2. List of heuristics based on [2]