User-Test Results Injection into Task-Based Design Process for the Assessment and Improvement of Both Usability and User Experience

Regina Bernhaupt[1], Philippe Palanque[2], François Manciet[1], and Célia Martinie[2(✉)]

[1] ruwido austria gmbh, Köstendorfer Straße 8, 5202 Neumarkt a. W., Austria
{regina.bernhaupt,francois.manciet}@ruwido.com
[2] ICS-IRIT, University of Toulouse, 118, route de Narbonne, 31062 Toulouse, France
{palanque,martinie}@irit.fr

Abstract. User Centered Design processes argue for user testing in order to assess and improve the quality of the interactive systems developed. The underlying belief is that the findings from user testing related to usability and user experience will inform the design of the interactive system in a relevant manner. Unfortunately reports from the industrial practice indicate that this is not straightforward and a lot of data gathered during user tests is hard to understand and exploit. This paper claims that injecting results from user-tests in user-tasks descriptions support the exploitation of user test results for designing the n+1 prototype. In order to do so, the paper proposes a set of extensions to current task description techniques and a process for systematically populating task models with data and analysis gathered during user testing. Beyond the already known advantages of task models, these enriched task models provide additional benefits in different phases of the development process. For instance, it is possible to go beyond standard task-model based performance evaluation exploiting real performance data from usability evaluation. Additionally, it also supports task-model based comparisons of two alternative systems. It can also support performance prediction and overall supports identification of usability problems and identifies shortcomings for user experience. The application of such a process is demonstrated on a case study from the interactive television domain.

Keywords: Task models · Usability evaluation · User experience evaluation · Process

1 Introduction

User-Centered Design (UCD) processes argue for user testing in order to assess and improve the quality of the interactive systems developed. User-centered design and development is typically performed iteratively with four major phases: (1) Analysis, (2) Design, (3) Development/Implementation and (4) Evaluation [4]. In the analysis phase the main goal is to understand who is (or will be) using the system, in what kind of environment and for what kind of activities or tasks. Many notations, processes and

tools have been proposed for gathering information about the users either in formal (via formal requirements as in [27] or formal task models [37]) or informal ways (via brainstorming [11] or prototyping [43]). One of the main advantages put forward by notations is that they make it possible to handle real-size applications and, if provided with a formal semantics, make it possible to reason about the models built with the notations and assess the presence or absence of properties.

The design phase encompasses activities to create or construct the system according to the analysis results. In the development phase the system is implemented and builds the basis for the evaluation of the interactive system [33].

One of the most used methods to evaluate early versions of a system or a prototype is user testing. The term user testing is broadly used in the area of human-computer interaction but in general describes any form of evaluation of an interactive system that involves users. User testing is associated with evaluating usability and now also incorporates user experience evaluation [18]. Goal of user testing is to gather feedback from users to identify usability problems or to understand what type of experience users have when interacting with the product. User testing can be classified into user tests that are either performed in the laboratory or in the field [29], where a user is asked to perform a set of tasks during that study. Users performance is typically observed and recorded (by video and for example measuring bio-physiological data) and users are asked for verbal responses (interviews) or ratings (e.g. via standardized questionnaires like SUS [8] or AttrakDiff [2].).

One key limitation for the majority of such user testing is how the results are made available for the next iteration in the design and development process [18]. To help inform the next iteration of a prototype, product and system, this information would be ideally fed back to the analysis phase to inform or improve design.

To support usability and user experience (UX) as key software qualities in the design and development cycle there is a need to:

(1) find a way to *document* the results of evaluation studies,
(2) allow the *comparison* of alternative approaches or systems,
(3) enable *prediction* of user behaviour and performance,
(4) support analysis of *usability* problems (efficiency, effectiveness and satisfaction), and
(5) show how different functions or tasks of the user contribute to the various user experience dimensions like aesthetics, emotion, identification, meaning and value and in general to the overall *user experience*.

Goal of this paper is to show a solution for how to integrate results from user tests in task models to support documentation, comparison, prediction and analysis of usability problems and relation of tasks to the overall user experience.

The paper is structured as follows: section two presents a state of the art on usability evaluation with a focus on user testing and an overview on user experience evaluation followed by a short overview on task analysis and modelling.

Section three describes our proposed process model and section four shows how this process model was applied in a case study in the area of interactive TV. We conclude with a summary and a discussion of the paper in section five.

2   State of the Art

2.1   Usability Evaluation Methods

Usability Evaluation is a phase in iterative UCD processes with the goal to investigate if the system is efficient to be used, effective when used and if users are satisfied [19]. Another central goal is to understand how users learn to use the system and (for complex systems) how to train users. Methods for usability evaluation can be classified in methods that are performed by usability experts, automatic methods, and those involving real end-users.

Usability evaluation methods that are performed by experts rely on ergonomic knowledge provided by guideline recommendations, or on the experts' own experiences to identify usability problems while inspecting the user interface. Known methods belonging to this category include Cognitive Walkthrough [23, 42], formative evaluation and heuristic evaluation [32] and benchmarking approaches covering issues such as ISO 9241 usability recommendations or conformance to guidelines [3]. Inspection methods can be applied in the early phases of the development process through analysis of mock-ups and prototypes. The lack of ergonomic knowledge available might explain why inspection methods have been less frequently employed. Automatic methods include approaches that enable automatic checking of guidelines for various properties (e.g. user interface design, accessibility…).

Methods involving real users are commonly referred to as user testing. User tests can be performed either in a laboratory or in the field with the main goal to observe and record users' activity while performing predefined activities that are typically described as scenarios [29] representing parts of the tasks users can perform with the overall system.

User tests are performance measurements to determine whether usability goals have been achieved. These measurements if performed scientifically rigorous are then called experiments or experimental evaluations [22], while tests with low numbers of participants and the main goal to identify usability problems are referred to as usability studies or usability tests [16].

A typical user test consists of several steps starting for example to obtain demographic information (e.g. gender, age, competencies, experiences with systems, …) and information on their preferences or habits. Participants provide that information typically by answering questionnaires or answering interview questions. A second step is then to ask users to perform a set of tasks. Their behaviour is observed and classified e.g. identifying if users performed the task successful, how long it took them to perform the task or how many errors were made. Users most often are video recorded (observation of certain behaviours, movements or reactions) and system interaction can be logged. Finally, users will provide feedback on the system e.g. filling out questionnaires or answering interview questions. Questionnaires have been extensively employed [40] to obtain quantitative and qualitative feedback from users (e.g. satisfaction, perceived utility of the system, user preferences for modality) [32] and cognitive workload (especially using the NASA-TLX method).

More recently, simulation and model-based checking of system specifications have been used to predict usability problems such as unreachable states of the systems or conflict detection of events required for fusion. [31] proposed to combine task models (based on Concur Task Tree (CTT) notation) with multiple data sources (e.g. eye-tracking data, video records) in order to better understand the user interaction.

2.2 User Experience and Its Evaluation

User Experience (UX) still misses a clear definition especially when it comes to the fact to try to measure the concept or related constructs or dimensions [21]. As of today the term user experience can be seen as an umbrella term used to stimulate research in HCI to focus on aspects which are beyond usability and its task-oriented instrumental values. UX is described as dynamic, time dependent [20] and beyond the instrumental [17]. From an HCI perspective the overall goal of UX is to understand the role of affect as an antecedent, a consequence and a mediator of technology. The concept of UX focuses rather on positive emotions and emotional outcomes such as joy, fun and pride [17].

There is a growing number of methods available to evaluate user experience in all stages of the development process. Surveys on these contributions are already available such as [5] who present an overview on UX and UX evaluation methods or HCI researchers who have summarized UX evaluation methods in a website [1]. Beyond that work on generic methods, contributions have been proposed for specific application domains, e.g. for interactive television [41]. User experience does include a look on all the (qualitative) experience a user is making while interacting with a product [28]. The current ISO definition on user experience focuses on a *"person's perception and the responses resulting from the use or anticipated use of a product, system, or service"* [19]. From a psychological perspective these responses are actively generated in a psychological evaluation process, and it has to be decided which concepts can best represent the psychological compartments to allow to measure the characteristics of user experience. It is necessary to under-stand, investigate and specify the dimensions or factors that are taken into account for the various application domains.

User experience evaluation is done in the majority of cases in combination with a usability study or test, applying additional UX questionnaires focusing on a selection of user experience dimensions. Examples are the AttrakDiff questionnaire [2] measuring hedonic and pragmatic quality and attractiveness, or Emo Cards [12] enabling the user to show their emotional state [1].

Data from user experience evaluation can be classified in qualitative (e.g. descriptions of feelings of a user when interacting with a system) or quantitative (e.g. rating scores). They can either reflect the user's experience for the whole system, or can be specifically associated to a task or sub-task (e.g. a physiological reaction like an increased heart-rate while doing a specific sub-task).

All these usability and user experience evaluation methods have a common limitation: they do not specify in detail how the evaluation results, for example reports of usability problems, task times, users' perception of difficulty for usability or appreciation levels, ratings or bio-physiological data for user experience can be used to inform the next design iteration.

2.3   Task Models: Benefits and Limitations

Introduced by [34, 37], tasks models for describing interactive systems are used during the early phases of the user-centered development cycle to gather information about users' activities. They bring several benefits when they are used throughout the development process and the operation time:

- They support the assessment of the effectiveness factor of usability as well as usability heuristic evaluation [10, 39];
- They support the assessment of task complexity [14, 33, 44];
- They support the construction of training material and training sessions [25];
- They support the construction of the documentation for users [15];
- They help to support the errors done by users as wells as their anticipations [13, 38];
- They help to identify the good candidates for migration [24, 45];
- They help to provide users contextual help [35, 36];
- They support the redesign of system [46].

   Nonetheless, task models suffer from various limitations:

- They miss quantitative information about performance data (number of errors per task, ratings for each task…);
- They miss connection to user experience and other software quality attributes;
- Tool support and process support is limited when it comes to the integration of usability and user experience evaluation data to inform the next iteration of design.

   In terms of tools there are only few available that allow to describe tasks not only representing activities but enabling the notion of error as well as the annotation of necessary knowledge and system used for the interaction [26]. We thus decided to extend the existing tool supported notation called HAMSTERS [13], as it is closest to what we would need for re-injecting results from user tests.

3   How to Enhance Task Models with Data: A Process Proposal

For any complex system that is developed following an iterative UCD process it has been reported that results from the usability evaluation phase of the system in stage (n) have not been incorporated in the next version of the prototype or system (n+1). We argue that task models can be beneficial in such an approach given that the tool support is able to represent an interactive system in detail.

   A task modelling tool thus has to be able to store the information gathered during user-tests related to usability and UX. It must allow to connect task descriptions with user test results to support the understanding and analysis of collected data related to the task models. This way it is possible to identify limitations and how small activities of the user, like performing a sub-task like a log-in to a system, can influence the overall perception of the user experience of the system. We have been choosing HAMSTERS as a tool to show how task models can be enhanced to show data gathered in user tests.

   We propose a **PR**ocess to **EN**hance **TA**sk **M**odels (PRENTAM) shown in Fig. 1, enabling the insertion of the data from the user tests in task models. Starting with (1) a

task analysis that is based on a variety of artefacts and insights obtained with methods like focus groups, interviews or ethnographic methods, the tasks a user can perform when interacting with the system are described. Based on the task analysis the tasks are modelled (2). Task modelling is supported by a variety of different tools; in our case HAMSTERS[1] [13].
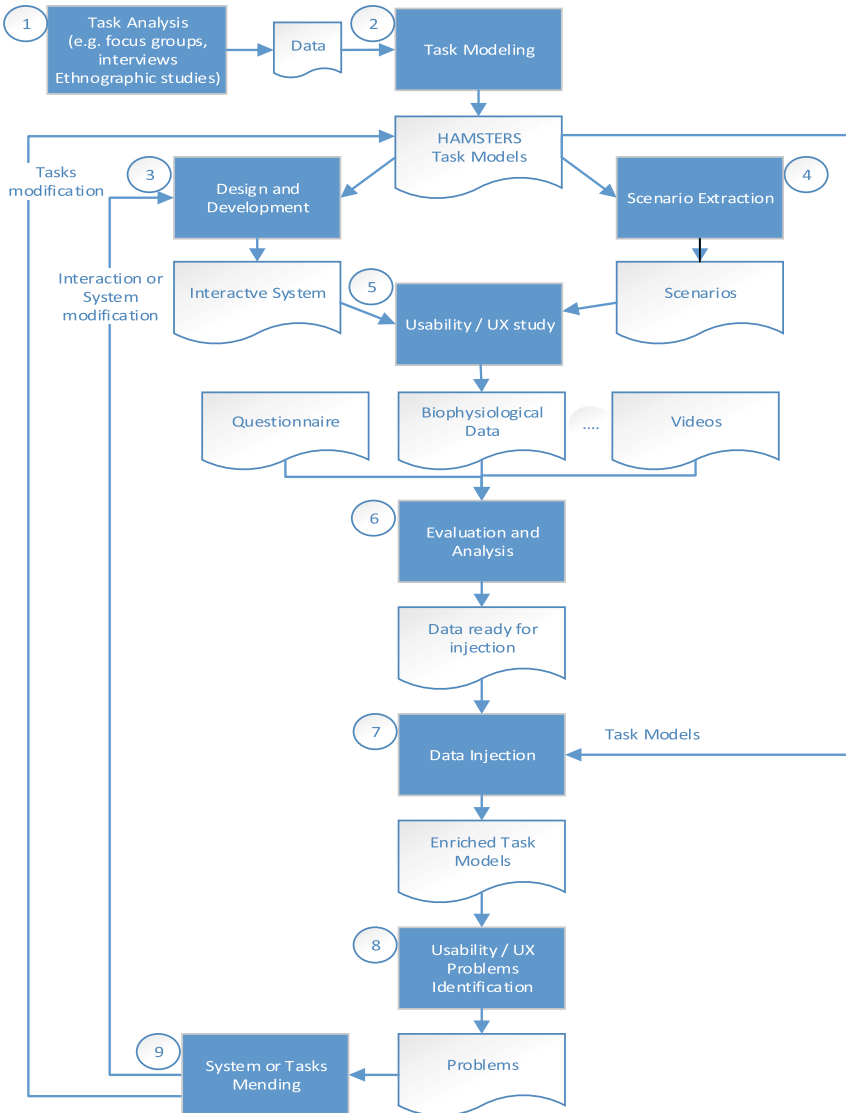

Fig. 1. Proposed process flow to enhance task models 'PRENTAM'

---

[1] https://www.irit.fr/recherches/ICS/softwares/hamsters/.

(3) Task models form the basis for the next step that is the design and development of the systems. We are aware that there is a variety of processes, methods and development stages included in this activity, but given that our contribution lies in how to re-inject user test results into task models, we just provide an abstract phase in the process model. Once a first prototype is available for evaluation the task models can be used to extract scenarios that shall be tested in the user test (4). The user test then is conducted through a Usability/UX study, following the same methods and procedures during the test for each of the participants (5). Within the user test each scenario is performed by a number of users. A set of different usability metrics can be measured which typically include metrics for effectiveness (successfully performed tasks, number of errors, task time in total, time measures for abstract tasks…) or satisfaction (users rating of the perceived difficulty). In terms of user experience dimensions are measured using questionnaires or interviews, observation like videos or eye-tracking, sometimes using bio-physiological feedback. Typical dimensions for user experience are aesthetics, emotion, identification, stimulation, meaning/value or social connectedness [7]. Data that is collected during such a user test thus can be (a) qualitative data like responses of a user in an interview, (b) quantitative data - ranging from ordeal to ratio data.

In (6), evaluation provides a multitude of data (sets) that has to be analysed, and where data is subsequently extracted. Analysis includes grouping of data (for example computing means and standard deviation) or statistical analysis (significance tests).

The analysed data then is injected in the task model (7). The important novelty aspect of this proposed process is that it enhances the task models used with the data gathered during the evaluation. Data is analysed and extracted from the evaluation and is injected inside the task models. Each task is enriched with data, for example minimal task time or maximal task time can be annotated as a property for the whole tasks or sub-tasks (see also later Fig. 3 for a depiction how this looks like in HAMSTERS).

Data is injected on several levels of the task model. For data that is related to the overall system evaluation this would be at the root of the task, while for data that is related to a small activity or an error this is at the nodes of a sub-task. The notation is using similar extensions as presented in [26].

The enriched task models can be used to understand and identify for example usability problems or limitations in terms of user experience (8). Depending on the problems found this will lead to changes in the system (System or Tasks Mending), e.g. by enhancing or improving an interaction technique or by re-ordering sub-tasks or activities (9). The enhanced task models then build the basis for the iteration of the system. In some cases, this can also lead back to the analysis phase.

This process ends when results from the user study are good enough to allow a release of the product. Given that changes are made to the system the new system has to be described starting again with a task analysis.

4   A Case Study from Interactive TV

Goal of this case study is to show how we followed the 'PRENTAM' during the development of an interactive TV system.

4.1 The Interactive TV Prototype

The prototypical system that we designed and developed enables the user to watch live television broadcast with associated functionalities including direct control like changing channels, regulating volume or muting the sound and additional functionality including Electronic Program Guide (EPG), a Video On Demand section (VOD), the support of personalization with individual user profiles, storage of personal data like photos and access to system settings (e.g. pin code registration to restrict access to content for children). The system also allows the user to control video content (forward, back, pause) and to time-shift programs. Figure 2 shows the main user interface and EPG. The system is based on a simple six button navigation (up, down, left, right, ok, back) with an overall good usability [40].
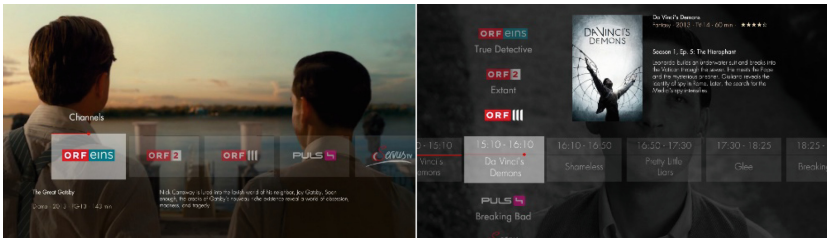

Fig. 2. Main menu (left) and electronic program guide (right) of the User Interface prototype.

In this Case Study we focus on one aspect of the system which was the introduction of the ability to transfer content from the TV to other devices via selectable menu options on the TV user interface. For example, this function allows the user to take away the movie being watched on the TV to a mobile device (e.g. a tablet).

4.2 Following the "PRENTAM" Process Step by Step

Task Analysis (Step 1). A task analysis was performed to understand activities and user goals that are related to using several devices while watching TV, including moving data or content (movies) between these devices. The task analysis involves approaches such as focus groups, interviews and ethnographic studies [41].

Task Modelling (Step 2). Based on the task analysis of the interactive TV system we modelled the main system tasks using HAMSTERS. This modelling was based on previous descriptions from [30].

Design and Development (Step 3). Based on the task models we have been designing and developing three additional functionalities: (1) enabling the user watching a TV show or program to access additional information related to that TV show on the tablet (2) allow users to take away the currently displayed TV show or program on the tablet and (3) to compare different movies in terms of user ratings before buying them.

Scenario Extraction (Step 4).   One advantage of using task models is the ability to use them to extract scenarios for the evaluation of the proposed system. We chose four scenarios that covered the three additional functionalities:

0. *User Test Scenario 0: Trial task to discover the system (change channels and access video on demand section);*
1. *User Test Scenario 1: Get more information (number of episodes) about a TV show in the program guide;*
2. *User Test Scenario 2: Get extra information about a movie (number of Os-car/ ratings);*
3. *User Test Scenario 3: Store information (ratings) about movies before buy-ing them in order to compare them;*
4. *User Test Scenario 4: Continue to watch a movie on the second screen.*

Usability and UX Study (Step 5).   Thirty-two students in computer science from the University of Toulouse took part in the study. Twenty-four were male and 8 were female. The age of participants ranged from eighteen to twenty-five, with an average of 21.7 (SD = 1.65).

The evaluation study took place in a room that was arranged with two sofas, one table and the desk where the TV screen and the audio system were placed. The TV screen used was in fact a 21.5″ computer screen, full HD. The second screen used was a tablet Google Nexus 7 running android 4.4. Users were video-taped during the session and we had an eye-tracker installed to follow the eye-gaze.

The evaluation study was structured into four parts. During the first one, we asked users questions about their media consumption habits, as well as their knowledge about TV systems and second screen apps. The second part was dedicated to the use of the system. The experimenter gave the user basic information about how the sys-tem works. Each participant conducted four tasks with the system. For each task, a short introduction into the scenario was given, followed by an explicitly formulated task assignment. Hints were provided after a predefined time period. Additionally, each task had a time limit. If a participant needed more time the task was stopped, counted as not solved and the correct way to solve the task was explained to the user. After performing the four tasks, and answering questions about each tasks (difficulty, comfort, naturalness of the inter-action technique) the experimenter asked the user to fill out the AttrakDiff [2] and the SUS [8] questionnaire. The final part was an inter-view and the debriefing of the user.

During the evaluation study the following types of data have been gathered:

(a) data about demographics (age, gender, media consumption habits)
(b) data about the use of the system including time needed (measuring time in the system) to complete a task, errors (performing a user observation with a written protocol by the test leader), general user behaviour (video), eye-gaze (Eye-Tracking) and user's appreciation of the system including ratings on user experi-ence like naturalness or comfort
(c) data related to the user experience (e.g. AttrakDiff) and usability (e.g. SUS) of the system
(d) interview data from the final interview.

Evaluation and Analysis (Step 6). After a first step of cleaning up the data by identi-fying outliers and verifying the video material, data was analysed and prepared for injection in the task model. This included a video analysis and re-visiting of observation protocols reporting number of errors, task success and failures, preparation of average task completion time, averages of ratings etc. (see Table 1).

Table 1. Tasks results: Average Execution Times for Successful Tasks (N, time, Standard Devi-ation) and Rating of Difficulty of the Task and Standard Deviation (SD).

| | Total | | | | |
	S	*H*	*F*	*Time (SD)*	*R (SD)*
T1	20	7	5	122 (51,53)	2,4 (1,21)
T2	25	6	1	111 (50,70)	2,1 (1,09)
T3	22	9	1	193 (68,26)	2,3 (1,02)
T4	39	2	1	67 (25,44)	1,6 (0,81)

For qualitative data like answers in interviews we summarized the number of positive and/or negative comments related to usability and identified comments related to six user experience dimensions including aesthetics, emotion, identification, stimulation, social connectedness and meaning/value [7].

Data Injection into Task Model (Step 7). Once the evaluation results are analysed and summarized the data related to the various usability and UX dimension is fed back into the task model. The tool HAMSTERS allows the users to describe properties of tasks and activities. Figure 3 shows screenshots of the tree tabs of the frame Properties associated to a task model in HAMSTERS tool. Second and third tabs of this frame are respectively presenting the Usability and the User Experience dimensions where data can be included.

For the different types of data there are various other ways to represent them in the task model. For data related to the overall system appreciation (AttrakDiff) or user ratings on the overall usability of the system (SUS) the data is stored at high levels nodes located at the top of the task tree. Data that is related to tasks or more specifically situa-tions e.g. when the user was reporting difficulties with the interaction technique at a special instance while performing a task are stored directly at the relevant node (typically a leaf of the tree).

Identification of Usability and User Experience Problems (Step 8). After the meas-urement information about usability and user experience has been entered in task prop-erties frame for each evaluated task in task models, usability and user experience issues can be analysed in an integrated way. The data can complement standard performance metrics like KLM [9]. Especially interesting in our case is an analysis of behaviours where performance times are rarely available like time to change a devices or time for speech and touch interaction for remote controls. In the presented case study, in terms of usability problem identification, the task models clearly showed that the tasks were rather long (see minimum and maximum execution time for sub-task "Transfer the displayed program to the tablet" in Fig. 3(a)), complicated (see field "Difficulty rating"
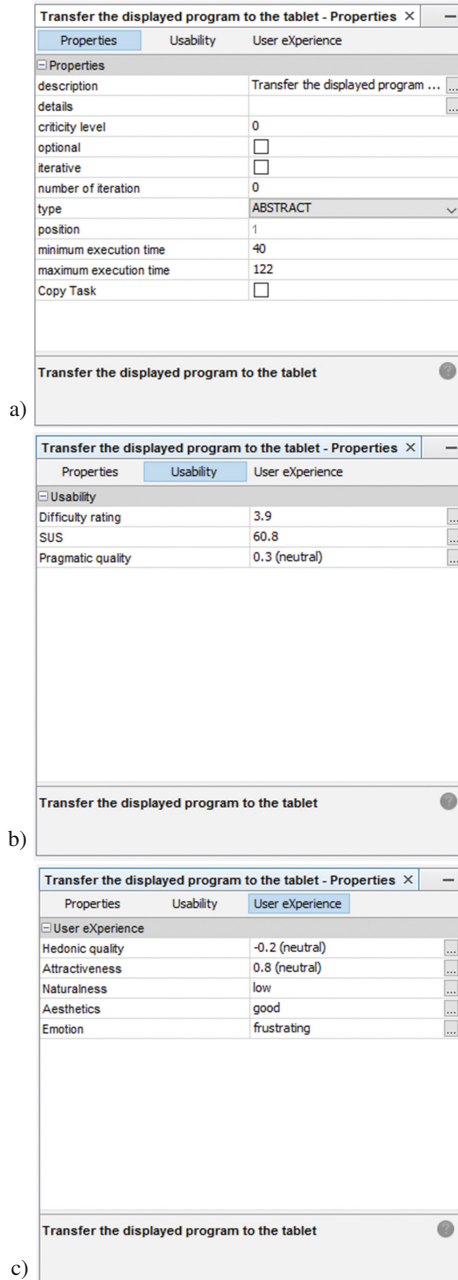
Fig. 3. In HAMSTERS tool, the Properties Frame Displaying Usability (b) and UX Data (c).

for the sub-task "Transfer the displayed program to the tablet" in Fig. 3(b)) and users were not much satisfied (see Fig. 3(b)), they would have preferred that the system performs these tasks automatically. In addition, in terms of user experience, the transfer if the viewed program from one device to another was perceived as less natural in terms of user experience and needed improvement.

Furthermore, when looking at user experience, the annotations in the task model allow the designer to revisit the task models and see what user experience dimensions are most important for the users for the different (sub-) tasks. They showed especially that users felt the tasks to be not natural (see field "Naturalness" for sub-task "Transfer the displayed program to the tablet" in Fig. 3(c)).

System or Tasks Mending (Step 9). To decrease the task difficulty for transferring the program to the second screen, as well as to improve the naturalness UX dimension of the interactive TV prototype, we decided to introduce an automation for these tasks were data is transferred between the devices. In order to enhance the usability and the user experience, the process of performing tasks including both a TV and a tablet have been simplified. By adding a remote control function on the tablet, the interactive TV prototype automatically communicates the current state/information from the TV to the tablet and supports the user accomplishing the task (e.g. take away of the movie). Figure 4 shows the new version of the task models for the task of content transfer including automation.

This new version of the task model is linked to a new version of the prototype. The insertion of results from user testing in task models was beneficial for the development of the system. Having task models for the iterations of the system allowed us to compare

Table 2. Differences between the two task models (Opt: number of optional tasks, Ite: number of iterative tasks, Nb: number of tasks)

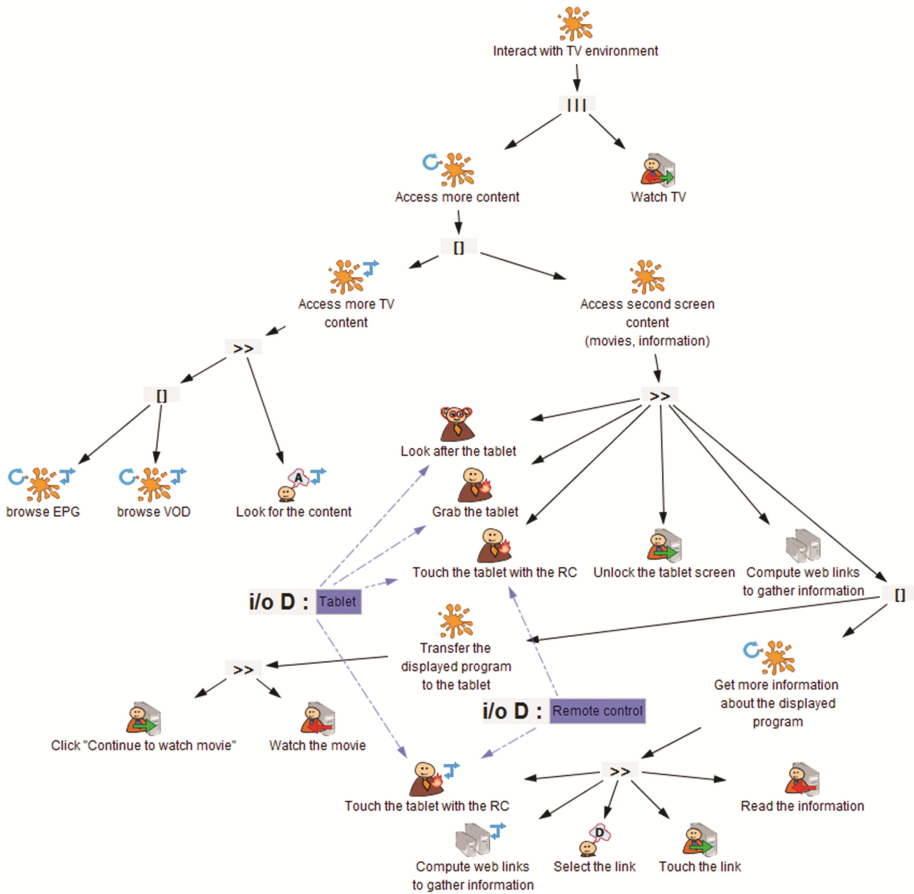|		Automation			No automation		
		Opt	*Ite*	*Nb*	*Opt*	*Ite*	*Nb*
User Tasks	*Perceptive Tasks*			1			1
	Motor Tasks	1	1	3			1
System Tasks	*System Tasks*	1	1	2			0
Interactive Tasks	*Input tasks*		1	3		2	7
	Output tasks		1	2		2	2
	I/O Tasks			0		1	1
Cognitive Tasks	*Cognitive Tasks*			0			2
	Cognitive Analysis Tasks			0			1
	Cognitive Decision Tasks		1	1		1	1
Total	***Total User Tasks***	*1*	*1*	*4*	*0*	*0*	*2*
	Total System Tasks	*1*	*1*	*2*	*0*	*0*	*0*
	Total Interactive Tasks	*0*	*2*	*5*	*0*	*5*	*10*
	Total Cognitive Tasks	*0*	*1*	*1*	*0*	*1*	*4*
	Total	***2***	***5***	***12***	***0***	***6***	***16***

Fig. 4. Task model: Showing an automated task that allows to transfer content by simply touching the device with the remote control to indicate the transfer.

how different types of automation affect the usability and the user experience of the system, and what changes in the tasks and sub-tasks provoke a change in the perception of the usability or user experience. Table 2 shows such an analysis of how the two task models are different, in terms of number of tasks and tasks type involved (optional tasks, or iterative tasks, i.e. tasks the user have to repeat several times).

Based on the evaluation of the automated system we found that efficiency and effectiveness were improved. For example, Usability was investigated using the SUS questionnaire. A closer inspection of the SUS scores revealed that the type of the system did have an observable influence on the SUS score (System A – with automation: mean = 83.2, SD = 13.0; System B - without automation: mean = 68.2, SD = 15.5). The results of the evaluation study have been published in [6].

5   Discussion: Benefits and Limitations

The interplay between usability evaluation and user interaction design is not as perfect as we would wish for [18]. In lots of cases, evaluation results are simply not taken into account for a design iteration, or are reinterpreted. Using a formal description including task models can help to improve such a feedback of evaluation results to be better (re) presented for design iterations. The proposed Task Model Enhancement Process (PRENTAM) supports design and development with the following:

(1) the selection of scenarios for usability and user experience studies (e.g. check of coverage)
(2) representation of evaluation data in the task model covering dimensions like user satisfaction that were not represented until now in a task model
(3) representation of user performance values in the task model to sup-port/complete predictive models like KLM
(4) validity checks if reality matches the assumptions and predictions. E.g. if the post completion error is really a problem: how often does the error happen and is that really an issue in terms of overall usability (and UX) perception
(5) the support for design and design decisions enabling to understand how to improve the design to overcome usability problems and user experience limitations, and to understand what parts of the current solution to keep (which avoids re-testing of these branches)
(6) predict (forecast) for new designs if they have the same structure (which limits the scope of the next evaluations)
(7) compare systems (e.g. different TV systems that support the same task can be compared) or compare interaction techniques (e.d. different types of interaction techniques for the same system).

In terms of user experience, the enhanced task model allows to understand how tasks do contribute to an overall UX judgement and the various dimensions of UX.

6   Summary and Conclusion

There is a fundamental belief when applying user-centered design and development processes that the findings from usability evaluations inform the user interaction design in a relevant manner. Unfortunately this is very often not the case [18]. To overcome this problem this article proposed the task model enhancement process (PRENTAM), that feeds back evaluation data into task models and enhances them. Applying this process to an (industrial) case study was a challenge but has shown that task modelling has its rightful place in a design and development cycle for large and complex interactive systems.

This is the first promising step toward systematic integration of usability and user experience evaluation data into artefacts used and produced in the design phases. The HAMSTERS tool currently only provides support to integrate quantitative values for usability and user experience.

Further contributions will be to integrate continuous data like observation data (videos, sound) or bio-physiological measurements (like eye-tracking, blood pressure…) which are currently simply connected via a link, but not integrated in the visual display of the task model.

We will continue to explore this Task Model Enhancement Process (PRENTAM) in the following iterations of the interactive TV prototype and at the same time start enhancing the HAMSTERS tool to support even more data visualization especially for the user experience dimensions.

References

1. All About UX: All UX Evaluation Methods. http://www.allaboutux.org/all-methods
2. AttrakDiff. www.attrakdiff.de
3. Bach, C., Scapin, D.: Ergonomic criteria adapted to human virtual environment interaction. In: Proceedings of the 15th Conference on L'Interaction Homme-Machine, pp. 24–31. ACM, New York (2003)
4. Baecker, R.M.: Readings in Human-computer Interaction: Toward the Year 2000. Morgan Kaufmann Publishers, San Francisco (1995)
5. Bargas-Avila, J.A., Hornbæk, K.: Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2689–2698. ACM, Vancouver (2011)
6. Bernhaupt, R., Manciet, F., Pirker, M.: User experience as a parameter to enhance automation acceptance: lessons from automating articulatory tasks. In: Proceedings of the 5th International Conference on Application and Theory of Automation in Command and Control Systems, pp. 140–150. ACM, New York (2015)
7. Bernhaupt, R., Pirker, M.: Evaluating user experience for interactive television: towards the development of a domain-specific user experience questionnaire. In: Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., Winckler, M. (eds.) INTERACT 2013, Part II. LNCS, vol. 8118, pp. 642–659. Springer, Heidelberg (2013)
8. Brooke, J.: SUS-A quick and dirty usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, A.L. (eds.) Usability Evaluation in Industry, pp. 189–194. Taylor & Francis, London (1996)
9. Card, S.K., Moran, T.P., Newell, A.: The keystroke-level model for user performance time with interactive systems. Commun. ACM **23**, 396–410 (1980)
10. Cockton, G., Woolrych, A.: Understanding inspection methods: lessons from an assessment of heuristic evaluation. In: Blandford, A., Vanderdonckt, J., Gray, P. (eds.) People and Computers XV—Interaction without Frontiers, pp. 171–191. Springer, London (2001)
11. Dennis, J.S.V.A.R.: Computer brainstorms: more heads are better than one. J. Appl. Psychol. **78**, 531–537 (1993)
12. Desmet, P., Overbeeke, K., Tax, S.: Designing products with added emotional value: development and application of an approach for research through design. Design J. **4**, 32–47 (2001)
13. Fahssi, R., Martinie, C., Palanque, P.: Enhanced task modelling for systematic identification and explicit representation of human errors. In: Abascal, J., Barbosa, S., Fetter, M., Gross, T., Palanque, P., Winckler, M. (eds.) INTERACT 2015, Part IV. LNCS, vol. 9299, pp. 192–212. Springer, Heidelberg (2015)

14. Fayollas, C., Martinie, C., Palanque, P., Deleris, Y., Fabre, J.-C., Navarre, D.: An approach for assessing the impact of dependability on usability: application to interactive cockpits. In: 2014 Tenth European Dependable Computing Conference (EDCC), pp. 198–209 (2014)
15. Gong, R., Elkerton, J.: Designing minimal documentation using a GOMS model: a usability evaluation of an engineering approach. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 99–107. ACM, Seattle (1990)
16. Gram, C., Cockton, G. (eds.): Design Principles for Interactive Software. Springer, Boston (1996)
17. Hassenzahl, M.: The interplay of beauty, goodness, and usability in interactive products. Hum.-Comput. Interact. **19**, 319–349 (2008)
18. Hornbaek, K., Stage, J.: The interplay between usability evaluation and user interaction design. Int. J. Hum.-Comput. Interact. **21**, 117–123 (2006)
19. ISO 9241-210 Ergonomics of Human-System Interaction Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems (2010)
20. Karapanos, E., Zimmerman, J., Forlizzi, J., Martens, J.-B.: Measuring the dynamics of remembered experience over time. Interact. Comput. **22**, 328–335 (2010)
21. Law, E.L.-C., Roto, V., Hassenzahl, M., Vermeeren, A.P.O.S., Kort, J.: Understanding, scoping and defining user experience: a survey approach. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 719–728. ACM, New York (2009)
22. Lazar, D.J., Feng, D.J.H., Hochheiser, D.H.: Research Methods in Human-Computer Interaction. Wiley, New York (2010)
23. Lewis, C., Polson, P.G., Wharton, C., Rieman, J.: Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 235–242. ACM, New York (1990)
24. Martinie, C., Palanque, P., Barboni, E., Ragosta, M.: Task-model based assessment of automation levels: application to space ground segments. In: 2011 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 3267–3273 (2011)
25. Martinie, C., Palanque, P., Navarre, D., Winckler, M., Poupart, E.: Model-based training: an approach supporting operability of critical interactive systems. In: Proceedings of the 3rd ACM SIGCHI Symposium on Engineering Interactive Computing Systems, pp. 53–62. ACM, New York (2011)
26. Martinie, C., Palanque, P., Ragosta, M., Fahssi, R.: Extending procedural task models by systematic explicit integration of objects, knowledge and information. In: Proceedings of the 31st European Conference on Cognitive Ergonomics, pp. 23:1–23:10. ACM, New York (2013)
27. Mavin, A., Maiden, N.: Determining socio-technical systems requirements: experiences with generating and walking through scenarios. In: Proceedings of the 11th IEEE International Requirements Engineering Conference, pp. 213–222 (2003)
28. McCarthy, J., Wright, P.: Technology as experience. Interactions **11**, 42–43 (2004)
29. McGrath, J.E.: Methodology matters: doing research in the behavioral and social sciences. In: Readings in Human–Computer Interaction, pp. 152–169. Elsevier (1995)
30. Mirlacher, T., Palanque, P., Bernhaupt, R.: Engineering animations in user interfaces. In: Proceedings of the 4th ACM SIGCHI Symposium on Engineering Interactive Computing Systems, pp. 111–120. ACM, New York (2012)
31. Mori, G., Paterno, F., Santoro, C.: CTTE: support for developing and analyzing task models for interactive system design. IEEE Trans. Softw. Eng. **28**, 797–813 (2002)
32. Nielsen, J., Mack, R.L. (eds.): Usability Inspection Methods. Wiley, New York (1994)
33. Norman, D.A., Draper, S.W.: User Centered System Design; New Perspectives on Human-Computer Interaction. L. Erlbaum Associates Inc., Hillsdale (1986)

34. Palanque, P.A., Bastide, R., Sengès, V.: Validating interactive system design through the verification of formal task and system models. In: Proceedings of the IFIP TC2/WG2.7 Working Conference on Engineering for Human-Computer Interaction, pp. 189–212. Chapman & Hall, Ltd., London (1996)
35. Palanque, P., Martinie, C.: Contextual help for supporting critical systems' operators: application to space ground segments. In: AAAI, pp. 7–11 (2011)
36. Pangoli, S., Paternó, F.: Automatic generation of task-oriented help. In: Proceedings of the 8th Annual ACM Symposium on User Interface and Software Technology, pp. 181–187. ACM, New York (1995)
37. Paternó, F., Mancini, C., Meniconi, S.: ConcurTaskTrees: a diagrammatic notation for specifying task models. In: Howard, S., Hammond, J., Lindgaard, G. (eds.) Human-Computer Interaction, INTERACT 1997, pp. 362–369. Springer, US (1997)
38. Paternó, F., Santoro, C.: Preventing user errors by systematic analysis of deviations from the system task model. Int. J. Hum.-Comput. Stud. **56**, 225–245 (2002)
39. Pinelle, D., Gutwin, C., Greenberg, S.: Task analysis for groupware usability evaluation: modeling shared-workspace tasks with the mechanics of collaboration. ACM Trans. Comput.-Hum. Interact. **10**, 281–311 (2003)
40. Pirker, M., Bernhaupt, R., Mirlacher, T.: Investigating usability and user experience as possible entry barriers for touch interaction in the living room. In: Proceedings of the 8th International Interactive Conference on Interactive TV&Video, pp. 145–154. ACM, New York (2010)
41. Pirker, M.M., Bernhaupt, R.: Measuring user experience in the living room: results from an ethnographically oriented field study indicating major evaluation factors. In: Proceedings of the 9th International Interactive Conference on Interactive Television, pp. 79–82. ACM, New York (2011)
42. Polson, P.G., Lewis, C., Rieman, J., Wharton, C.: Cognitive walkthroughs: a method for theory-based evaluation of user interfaces. Int. J. Man Mach. Stud. **36**, 741–773 (1992)
43. Rettig, M.: Prototyping for tiny fingers. Commun. ACM **37**, 21–27 (1994)
44. Swearngin, A., Cohen, M.B., John, B.E., Bellamy, R.K.E.: Human performance regression testing. In: Proceedings of the 2013 International Conference on Software Engineering, pp. 152–161. IEEE Press, Piscataway (2013)
45. van Welie, M., van der Veer, G.C.: Groupware task analysis. In: Handbook of Cognitive Task Design, LEA, NJ, pp. 447–476 (2003)
46. Wilson, S., Johnson, P.: Bridging the generation gap: from work tasks to user interface designs. In: Computer-Aided Design of User Interfaces I, Proceedings of the Second International Workshop on Computer-Aided Design of User Interfaces, Namur, Belgium, 5–7 June, pp. 77–94 (1996)