Multi-Dimensional Context-Aware Adaptation of Service Front-Ends

Project no. FP7 – ICT – 258030

Deliverable D.5.3.1
First Evaluation
(users, development tools)

Due date of deliverable: 30/09/2012

Actual submission to EC date: 30/09/2012

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination level		
[PU]	**[Pubic]**	**Yes**

Document Information	
Lead Contractor	TID
Editor	ISTI-CNR
Revision	v.1.0 (30/09/2012)
Reviewer	TID
Approved by	TID
Project Officer	Michel Lacroix

Contributors	
Partner	Contributors
ISTI	Fabio Paternò, Carmen Santoro, Lucio Davide Spano,
SAP	Jörg Rett, Yucheng Jin, Sara Bongartz
W4	Nicolas Bodin, Jean-Loup Coméliau
TID	F. Javier Caminero-Gil

Changes			
Version	Date	Author	Comments
1	12/09/2012	ISTI	First document draft
2	20/09/2012	ISTI	First consolidated version
3	28/09/2012	W4/TID/SAP	Content update
4	28/09/2012	ISTI	Final version

Executive Summary

This report describes the results of some initial prototype evaluations that have been conducted within the Serenoa Project in order to assess the considered adaptations.

In particular, it reports the evaluation carried out at SAP on the adaptive HMD-based prototype, the test conducted by CNR to assess the adaptation of a multimodal UI in the car rental domain, the work done at W4 on evaluation of a prototype in a business scenario, and the evaluation carried out by TID on two adaptive pilots (HealthDrive and SARA).

Table of Contents

1 Introduction

1.1 Objectives

The objective of this document is to report on a first evaluation of prototypes developed in Serenoa.

1.2 Audience

Being a public deliverable, this document will be available outside the confines of the project's consortium and is intended to be of interest to the following parties:

a) Members of the consortium, who will find here a detailed description of the fundamentals of the Description Language that the project is to use in the future.

b) Researchers in the relevant fields: adaptation of SFEs, UI theorists, descriptive languages and medium-scale project software engineering.

c) EC officials that will use the information in this document as an account of the activities taken in the project tasks that inform this work.

1.3 Related documents

- **D2.4.1 Criteria for the Evaluation of CAA of SFEs (R1)** indicated a first set of evaluation criteria that can be relevant for the project.
- **D2.4.2 Criteria for the Evaluation of CAA of SFEs (R2)** aims to provide a self-contained update of D2.4.1 (R1), by indicating a revised set of evaluation criteria relevant for the project. It provided a set of criteria actually used to evaluate the Serenoa prototypes, as shown in this document D5.3.1.
- **D5.2.2 Application Prototypes (R2)** includes the second release of the application prototypes.

1.4 Organization of this document

Section 1 describes the scope and the organization of this document. Section 2 describes the HMD evaluation conducted at SAP. Section 3 presents first results of the evaluation on a multimodal application on car rental domain conducted by CNR. Section 4 presents W4's work on evaluation of a prototype in a business domain. Section 5 presents first results of TID evaluation on two adaptive prototypes. Section 6 presents the conclusions of this document and the planned future work.

2 HMD Evaluation at SAP

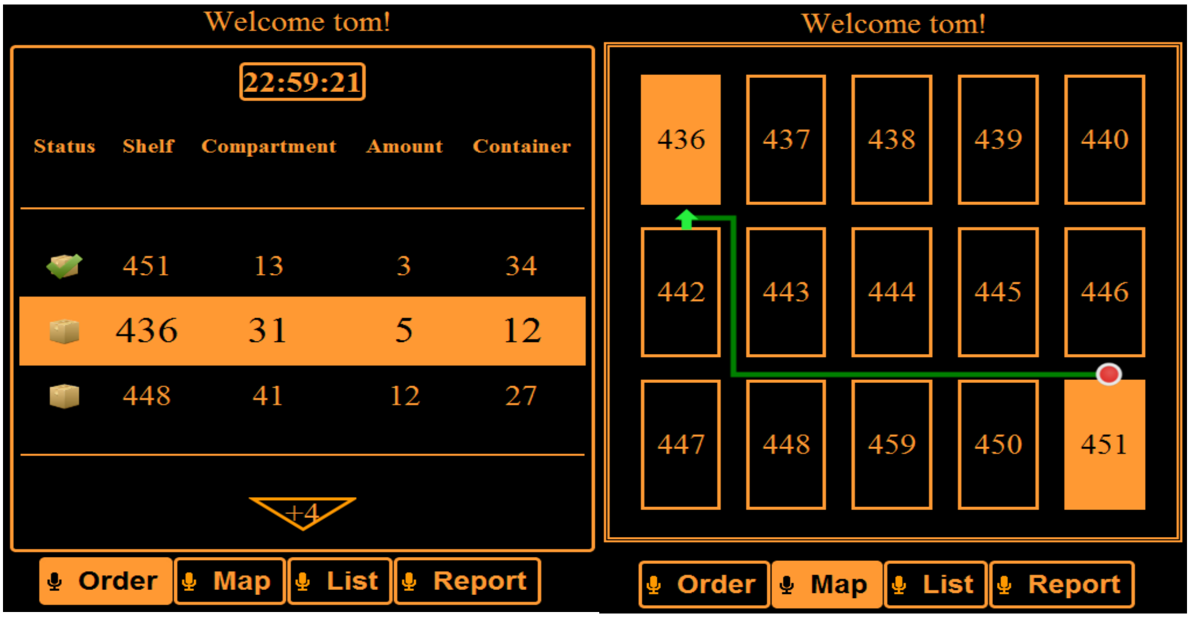
In this section we present a prototype of an adaptive warehouse order picking system consisting of an adaptive, context-sensitive UI which is based on an architecture for context-sensitive service front-ends. The details on the architecture have been described (Bongartz et al., 2012), and they have been evaluated in different phases according to the principles of UCE. Based on a first user study result of a low-fidelity prototype, we extracted usability problems specific to the adaptive features of the application. After that, we conducted a second user study with an improved high-fidelity prototype. Finally, we draw some conclusions regarding the design of Abstract User Interfaces (AUIs) and provide indications for future work.

2.1 The HMD Evaluation Adaptive Prototype

2.1.1 The Adaptive Prototype

Warehouse picking is a part of a logistics process often found in retail and manufacturing industries. The adaptive application presented here is enhanced with context aware features, which consider user-related aspects (tasks to accomplish, personal preferences and knowledge, etc.), technical aspects (available interaction resources, connectivity support, etc.) and environmental aspects (level of noise, light, etc.).

The graphical user interface (GUI) consists of four views (Order, Map, Task and Report). For the sake of brevity only the Order view and the Map view are discussed here. The Order view (shown in Fig.1) mainly contains information on the previous (i.e. shelf 451), the current (i.e. shelf 436) and the next (i.e. shelf 448) items to be picked. This sequence of picks is represented in three rows starting with the previous pick and having the current pick highlighted (i.e. inverting the foreground and the background colour) and magnified.



Welcome tom!

22:59:21

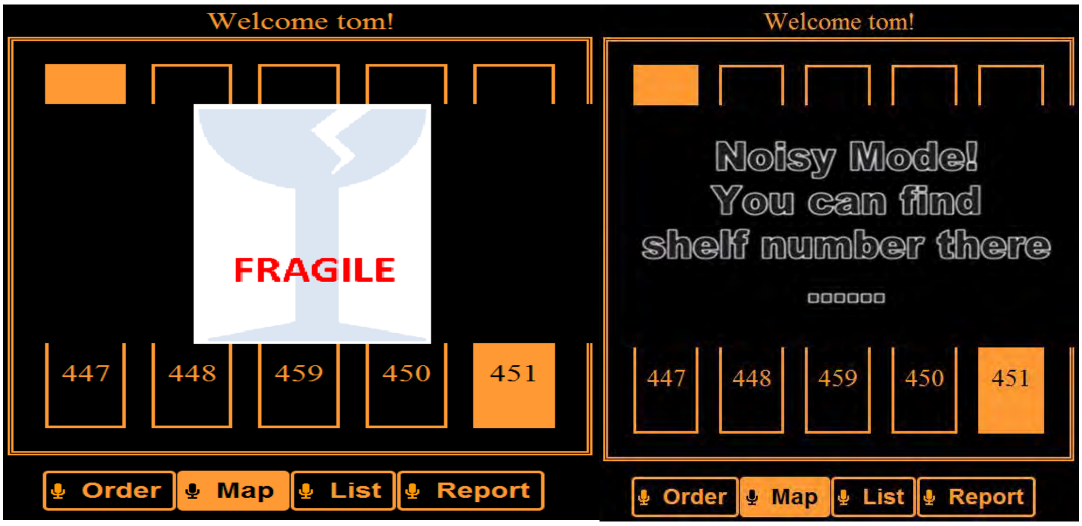
Status	Shelf	Compartment	Amount	Container
	451	13	3	34
	436	31	5	12
	448	41	12	27

+4

Welcome tom!

436	437	438	439	440
442	443	444	445	446
447	448	459	450	451

a) b)



Welcome tom!

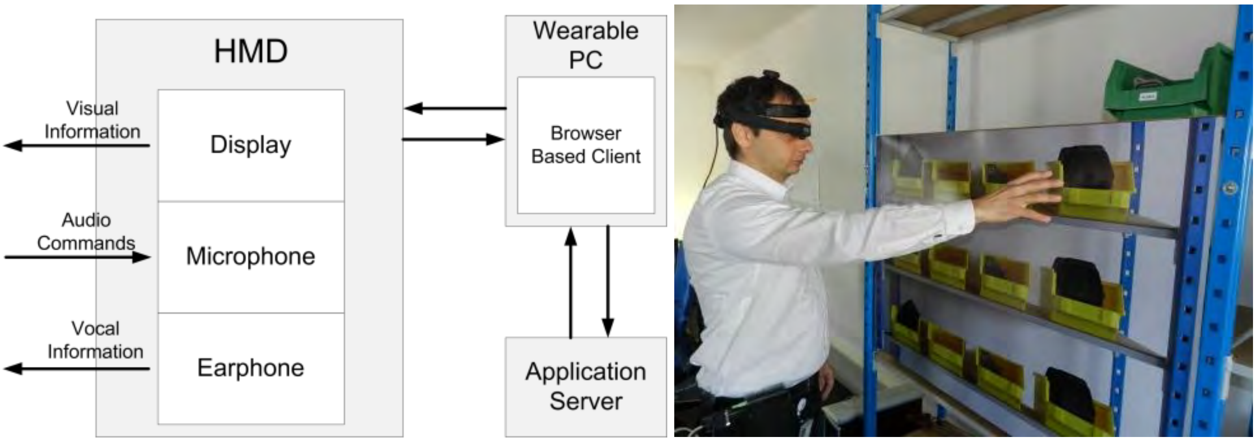
Noisy Mode! You can find shelf number there

c) d)

Fig. 1.: Design of the graphical user interface (GUI). a) Order view b) Map view c) User support in fragile mode d) User support in noisy mode

The columns reflect the types of information available for the pick (status, shelf, compartment, amount and container), while only the status of the pick (e.g. open), the shelf identifier (e.g. 473) and the amount of items to be picked (e.g. 7) are relevant here. The active view is reflected as a highlighted tab in the bottom area. The main information in the Map view is a simplified representation of the location of the shelves (in Bird eyes view) showing the current location of the picker (i.e. the previous shelf), the destination shelf (i.e. 473) and a suggested route (green line with arrow and red start point). Users can switch between the four views by speaking the name of the respective tabs.

A Head-Mounted Display (HMD) and a wearable computer are used to access the application. The UIs are implemented in HTML5, JavaScript and AJAX. The navigation route in the Map view is drawn using an HTML5 canvas. Speech recognition is realized using the Google speech recognition engine. The architecture of the application implementation is shown in Fig. 2. The display is used for the visual output, the earphone for the vocal output and the microphone for the vocal input of the user.



a) b)

Fig. 2. a) Architecture of the prototype. b) Picking from a shelf using a Head-Mounted Display

The basic interaction sequence (i.e. the basic interaction flow) with an example for an adaption is shown in Fig. 3: the picker is presented with three screens and two vocal outputs (upper balloons) and needs to perform two vocal inputs (lower balloons). Assuming that a picker who is experienced, i.e. has been working for a long time in the warehouse environment and thus should know by heart the location of the shelves, the Map view can be omitted. We assume that an indicator of the experience level is stored within the profile of the picker and is added as context information at run-time during the log-in procedure.

Table 1 lists the five variations of the context and its consequences for the interaction modalities with respect to the basic interaction flow. The adaptation server sends the updated data to the wearable computer after a change in the context has triggered the execution of an adaptation rule. Some changes might be triggered by the smart environment (e.g. tracking of the picker's position or the item's location).

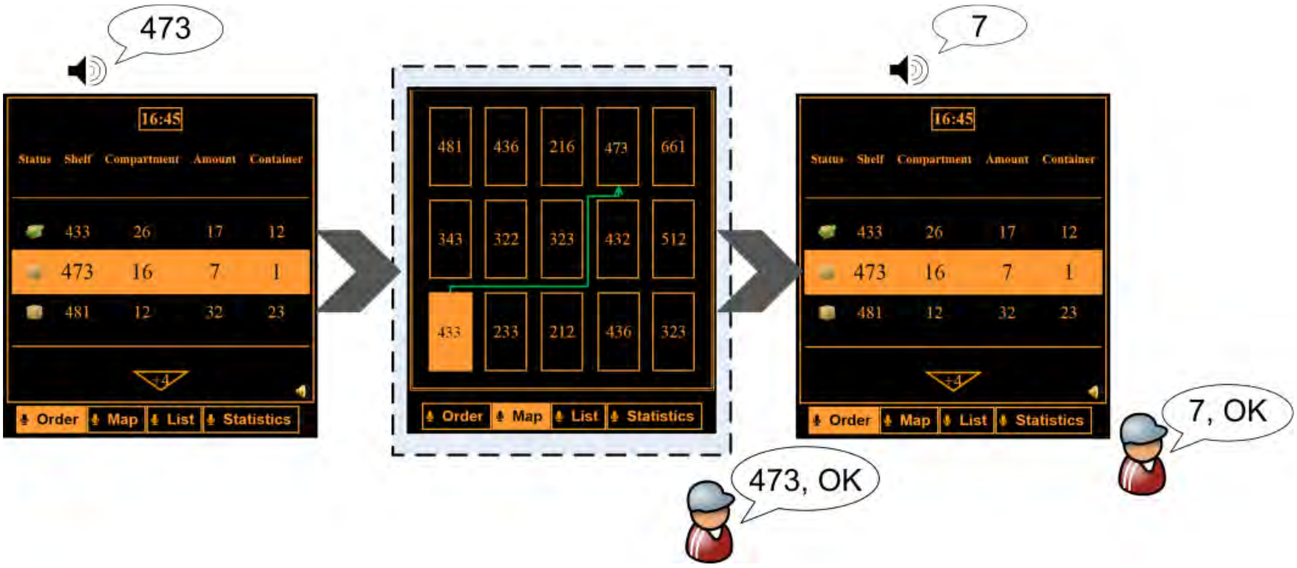

473
16:45
Status Shelf Compartment Amount Container
433 26 17 12
473 16 7 1
481 12 32 23
Order Map List Statistics
481 436 216 473 661
343 322 323 432 512
433 233 212 436 323
Order Map List Statistics
7
16:45
Status Shelf Compartment Amount Container
433 26 17 12
473 16 7 1
481 12 32 23
Order Map List Statistics
473, OK
7, OK

Fig. 3. Basic interaction flow with adaptation: the execution of the rule for an experienced picker omits the appearance of the Map view (dotted line)

2.1.2 User-Centered Evaluation

Following the principles of UCE in the design process of our AUI, we conducted two evaluations, one with a low-fidelity and a high-fidelity prototype each. Addressing usability problems found in the first study, the second study was aimed at evaluating the effect of subsequent improvements on the prototype.

In order make both studies statistically and conceptually comparable, we use the same questionnaires and study design in both studies. We present and compare the results of the two user studies and draw conclusions regarding the design of AUIs.

Context variation	Interaction consequence
The items to be picked are fragile	After vocally confirming the arrival at the destination by the picker, the visual output will be switched off, only vocal remains.
The route is blocked by other pickers	The Map view marks the blocked path and suggests an alternative route.
The picker is experienced	The Map view is omitted.
The environment is noisy	The vocal input and output is switched off, only visual output remains
The picking is not performed due to some confusion or distraction	An image of the item to be picked is shown, the vocal output is repeated.

Table 1. Variations of the context and its consequences for the interaction modalities

2.1.3 User Study 1

We have conducted a first user study in order to evaluate the five adaptation rules from the end-users point-of view (see (Bongartz et al., 2012)). The study aimed at evaluating the applicability and usefulness of the adaptation rules by assessing the quality of the adaptation rules as subjectively perceived by the participants. The general concept "quality" was operationalized by several more specific constructs, e.g. usefulness, comprehensibility or simplicity, which were assessed by a questionnaire.

To address such issues, the five adaptation rules were the independent variables. We had a within-subject design, meaning that every participant was confronted with every adaptation rule. The dependent variables were the subjectively perceived quality of the adaptation rule as assessed in a 9-item questionnaire. The questions originated from a list of non-functional requirements for the prototype, which were identified in user studies at the beginning of the project and aimed at assessing the following aspects: the user's awareness for the adaptation rule, its appropriateness and comprehensibility, its effectiveness with respect to performance and usability, its error-prevention, continuity, intuitiveness, and general likeability.

Participants were company staff or students of the local university. A total of 10 participants took part in the study, 9 were male and 1 was female. The average age of participants was 24 years (SD = 1.82). The technical set-up consisted of an HMD with earphone worn by the participants. The device presented the GUI and the vocal output as shown before. The sequence of the interaction was controlled by the moderator simulating the change of context and the execution of the adaptation rule.

Participants were first introduced into the scenario and the interface, i.e. getting familiar with the hypothetical situation in the warehouse and learning how to interact with the interface. Participants were asked to play through a "basic interaction flow" which started with the systems request to pick items from a certain shelf, required the user to hypothetically walk to that shelf and ended with the user's confirmation that he picked a certain amount of items. Participants were asked to comment their hypothetical actions, e.g. by saying "I walk to the shelf 473 now" or "I pick 7 items from the shelf". After ensuring that the participants understood the basic interaction flow of the interface, the study started by introducing the first alternative flow. All alternative flows (flows containing adaptation rules) were applied to the same scenario as practiced in the basic flow. Prior to playing through the alternative flows, participants were informed about the condition of the adaptation rule (e.g. "imagine you are now in a noisy environment"), but not about the actual rule (i.e. the action of the rule). All five rules were played through and the sequence of the adaptation rules was permutated to avoid order effects. After each rule, the 9-item questionnaire was filled out.

Since most of the scales of the questionnaire were not normal-distributed, we applied non-parametric tests for the data analysis. We calculated the Friedman test for every single questionnaire scale and the aggregated overall rating from all 9 scales (Bonferroni-corrected) to assess differences between the five adaptation rules. In case of significance, we calculated a post-hoc Wilcoxon signed-rank test for each pair of adaptation rule (Bonferroni-corrected as well).

The Friedman test revealed significant differences for the aggregated overall rating over all 9 scales ($\chi^2(4)$ = 18.74, p = .001) and for 4 of the subscales: Appropriateness ($\chi^2(4)$ = 19.26, p = .001), Performance (Z = -2.69, p=.007), Error-Prevention ($\chi^2(4)$ = 22.73, p = .000), Intuitiveness ($\chi^2(4)$ = 22.31, p = .000) and General Likeability ($\chi^2(4)$ = 18.92, p = .001). Only these significantly different scales are regarded in detail here. Post-hoc tests revealed a significant difference in the rating between the rules Fragile Objects and Traffic Jam (Z = -2.60, p = .009) and Experienced Worker and Traffic Jam (Z = -2.70, p=.007). The significant differences in the subscale Appropriateness are between the rules Fragile Objects and Traffic Jam (Z = -2.62, p = .009) and Fragile Objects and Pick Timeout (Z = -2.69, p = .007). For the subscale Error prevention, the significant differences can be found between the rules Fragile Object and Pick Timeout (Z = -2.71, p = .007), Traffic Jam and Experienced Worker (Z = -2.81, p = .005) and Pick Timeout and Experienced Worker (Z = -2.68, p = .007). Intuitiveness shows significantly different values for the rules Fragile Objects and Traffic Jam (Z = -2.69, p = .007). Finally, although the Friedman test revealed significant differences between the rules for the scales: general Likeability and Performance; direct pairwise comparison failed reaching significance due to Bonferroni correction.

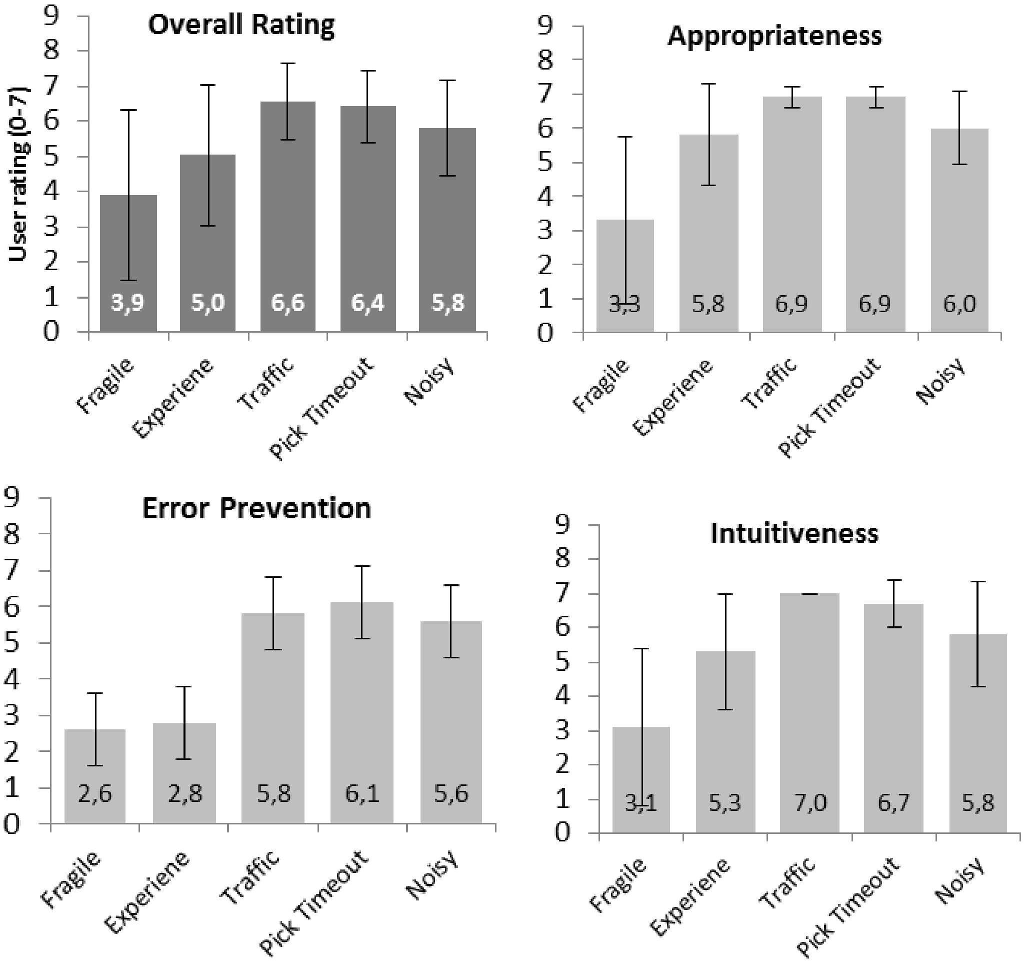


Fig. 4. Study 1: Overall rating and the subscales Appropriateness, Error-Prevention and Intuitiveness

The big picture of the results (see Fig. 4) shows a clear trend: all quality aspects of the Fragile Object rule are consistently rated the worst, and the Traffic Jam and Pick Timeout rule are consistently rated best. This pattern can be observed for all quality scales, indicating a clear and coherent preference pattern. Traffic Jam and Pick Timeout are consistently and undoubtedly preferred by the users (with very good overall ratings of 6.6 and 6.4 on a scale from 0-7). Alongside the good rating of these two rules, the standard deviation is very small, indicating a very high agreement between the participants. However, the Fragile Object rule, as the worst rated one, shows the highest variance in the ratings between the subjects. This indicates that there is no strong agreement between the subjects, yet still most of the subjects gave comparably low ratings for that rule. A possible explanation for this finding can be drawn from the subject’s comments. While all subjects gave a positive opinion about the idea to support the process of picking a fragile object, most of the subjects noted that the actual realisation of that rule was poor. Turning off the display was irritating and non-intuitive to the subjects. The abrupt darkness in the HMD was perceived as a break-down of the system and therefore caused confusion. Rather, subjects had wished to receive a short warning message before turning off the display.

We found similarities between those rules that were ranked well and those that were ranked poor. The group of poorly ranked rules was omitting information like the visual output and the Map view with regard to the Basic Interaction Flow. The Fragile rule takes a prominent position as a very strong modality, the visual channel, is shut off. Those rules that were ranked well however delivered additional information like the blocked path or the image of the item. This noticeable difference between the adaptation rules is presumably the reason for the striking difference in the preference ratings. Therefore, in the second study, we investigated the role of adding vs. removing information in the course of interface adaptation. The second study tested the hypothesis that the poorly ranked adaptation rules will be higher ranked when information is not only removed but the removal of information is actually explained beforehand by adding information.

2.1.4 User Study 2

The goal of the user study 2 was to evaluate whether the comparably poor performance of the rules Fragile Object, Experience User and Noisy Environment was improved by adding information (i.e. also called user support in (Paymans et al., 2004) prior to showing the adaptation in UIs. User support means the forgoing explanation of an occurring adaptation or hints of an approaching adaptation. The design of the study is same as in user study 1. However, in user study 1, we used a paper based map to simulate the warehouse layout and in user study 2 we simulated the warehouse environment on the ground of a huge meeting room, having papers as shelves and real items on the shelves representing the items to be picked (see Fig. 5). Consequently, users were truly able to move around and pick the items, which made the setting more realistic. The conditions for the adaptation rules were also implemented in a more realistic way, e.g. by putting obstacles in the way for the Traffic Jam rule or using real fragile objects (glasses) for the Fragile rule. An alongside research question was therefore, if the more realistic setting affects the evaluation results. This means, since 2 out of four rules (Traffic Jam and Pick Timeout) were not changed, the more realistic setting of the second study would not affect the reliability of evaluation if the evaluation scores of these two rules did not change.

Participants again were company staff or students of the local university (who did not participate in the first study). A total of 10 participants took part in the study, 9 were male and 1 was female. The average age of participants was 29 years (SD = 4.44).

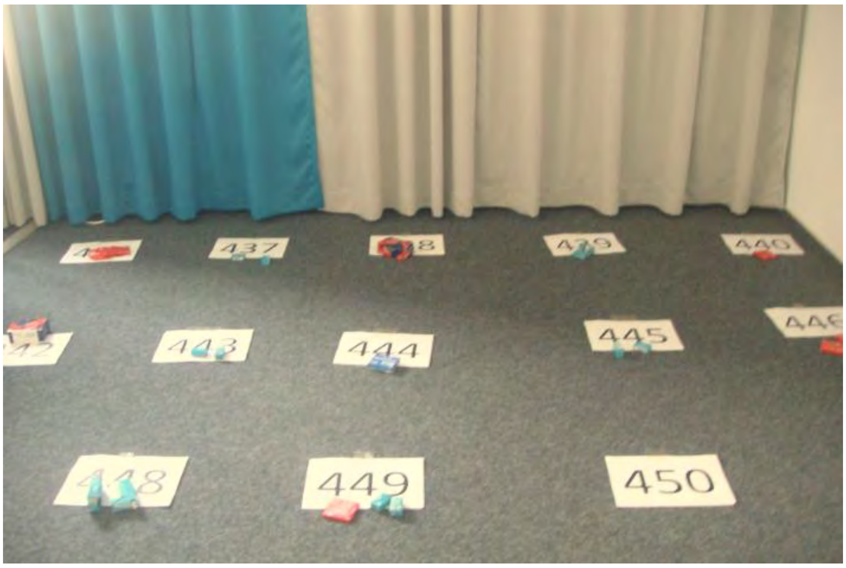


Fig.5 Evaluation Environment of User Study 2

Since most of the scales of the questionnaire were not normal-distributed, we applied non-parametric tests for the data analysis. We calculated the Friedman test for every single questionnaire scale and the aggregated overall rating from all nine scales (Bonferroni-corrected) to assess differences between the five adaptation rules. In case of significance, we calculated a post-hoc Wilcoxon signed-rank test for each pair of adaptation rule (Bonferroni-corrected as well).

The Friedman test revealed significant differences for the aggregated overall rating over all 9 scales ($\chi^2(4)$ = 17.99, p = .001) and for three of the subscales: Error-Prevention(χ^2 (4) = 17.76, p = .001), Intuitiveness (χ^2 (4)= -17.19, p=.002) and User Experience (χ^2 (4) = 15.96, p = .003). The scales with significant differences between the rules are displayed in Fig. 6. Although the Friedman test revealed significant differences between the rules for all these scales; pairwise comparison failed reaching significance due to Bonferroni correction. Taking a look at the graphs, there are three main interesting observations:

- The Fragile rule improved significantly compared to the first study.
- The Experiences Worker rule performs consistently worse than the other rules (although pairwise

comparison did not reach significance).

- The four other rules Experienced Worker, Traffic Jam, Pick Timeout and Noisy did not change in the course of the second experiment

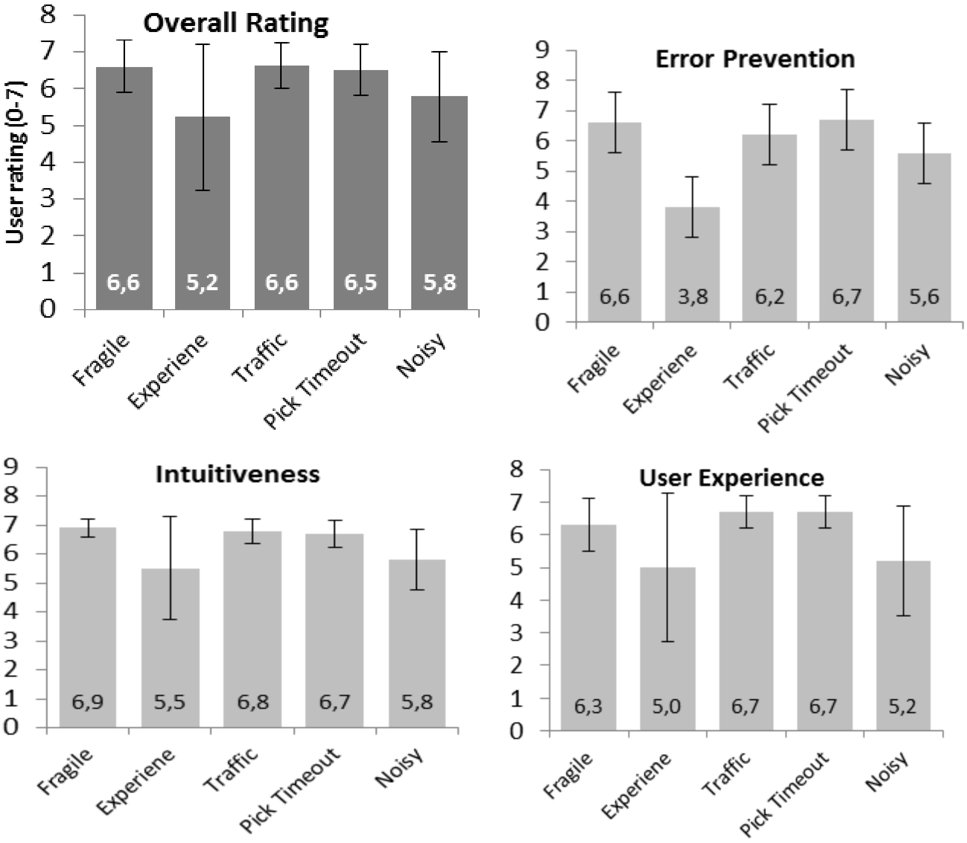


Fig. 6 Study 2: Overall rating and the subscales User Experience, Error-Prevention and Intuitiveness

In order to test these observations for significance, we conducted a Kruskal-Wallis-Test comparing the results of the first and the second study. The test reveals that the Overall Rating of the Fragile rule increased significantly (H(1) = 12.17, p=.000), which can be attributed to the scales Appropriateness (H(1) = 9.44, p = .002), Performance (H(1) = 11.14, p = .001), Error Prevention (H(1) = 11.44, p = .001), User Experience (XX(1) = 7.15, p = .008), Intuitiveness (XX(1) = 12.75, p = .000) and general Likeability (XX(1) = 8.07, p = .005). Thus, for the Fraglie rule, all scales except Continuity and Comprehensibility increased significantly. All other comparisons were not significant. Thus, all other rules were not rated better or worse (for no scale) compared to study 1.

2.2 Discussion

User study 2 addressed the research question: does the addition of information prior to the removal of information in the course of an adaptation of the interface improve the perceived quality of the adaptation rule? The results of the study partly support this hypothesis. While the Fragile rule was improved significantly in almost all the scales, the Experience and Noisy rules did not improve.

The improvement of the Fragile rule can most probably be attributed to what Paymans et al. (2004) call user support. According to the authors, users experience difficulty in building adequate models of adaptive systems, therefore user support is expected to help users understand and learn the adaptive rules. For the Fragile rule, the performance improved significantly with the help of user support. Before shutting down the display of HMD, the users have been notified by a short alert video to be cautious for picking fragile objects, so the rational of the rule can be more easily understood (prevent the user from visual distraction).

However, for the Experienced Worker and Noisy rule, the ratings are not improved by adding explanatory information as user support. We can think of two possible reasons for this finding. First, autonomous interface adaptations can easily reduce the usability of a system. Loss of control might be an issue in both rules. For example in the Noisy rule, users cannot confirm their location or the amount number by voice; instead the system will set a timeout for automatic confirmation. Setting the timeout either too long or too short will consequently put the user in an uncomfortable situation (i.e. waiting for or missing the following system information). In the Experienced Worker rule, the user might want to decide himself if he gets to see the map or not; although he might not really need it. In both cases, the loss of control over the system might be a problem. To overcome the problem of controllability, we can enrich the user profile and context information to provide even more precise and personalized adaptations. Furthermore, we can also consider increasing the flexibility of operation, so that users have more possibilities to intervene the adaptation. Second, even in user study 2, the setting of the user study is still simulated. A real testing environment with real users (i.e. real pickers) might result in different ratings. Although the change in the fidelity between the two studies presented here did not affect the ratings (see below); a real environment with real users might yield to more valid results (e.g. to imagine being an experienced user might not result in the same rating as actually being an experienced user).

Furthermore, the change in the evaluative setting did not affect the rating of the rules. This is an interesting finding with regard to evaluation methodologies. Although the study design was much more realistic in the second study, the ratings of the unchanged rules Traffic Jam and Pick Timeout were exactly the same for both studies. Thus we can conclude that a low-fidelity evaluation setting (e.g. imagining the movement through a warehouse vs. actually moving through a simulated warehouse) does not affect the fidelity of the ratings when evaluating adaptive features of an interface. Our studies suggest that the rating of adaptive rules has no direct and obvious relation to the fidelity of the evaluation environment.

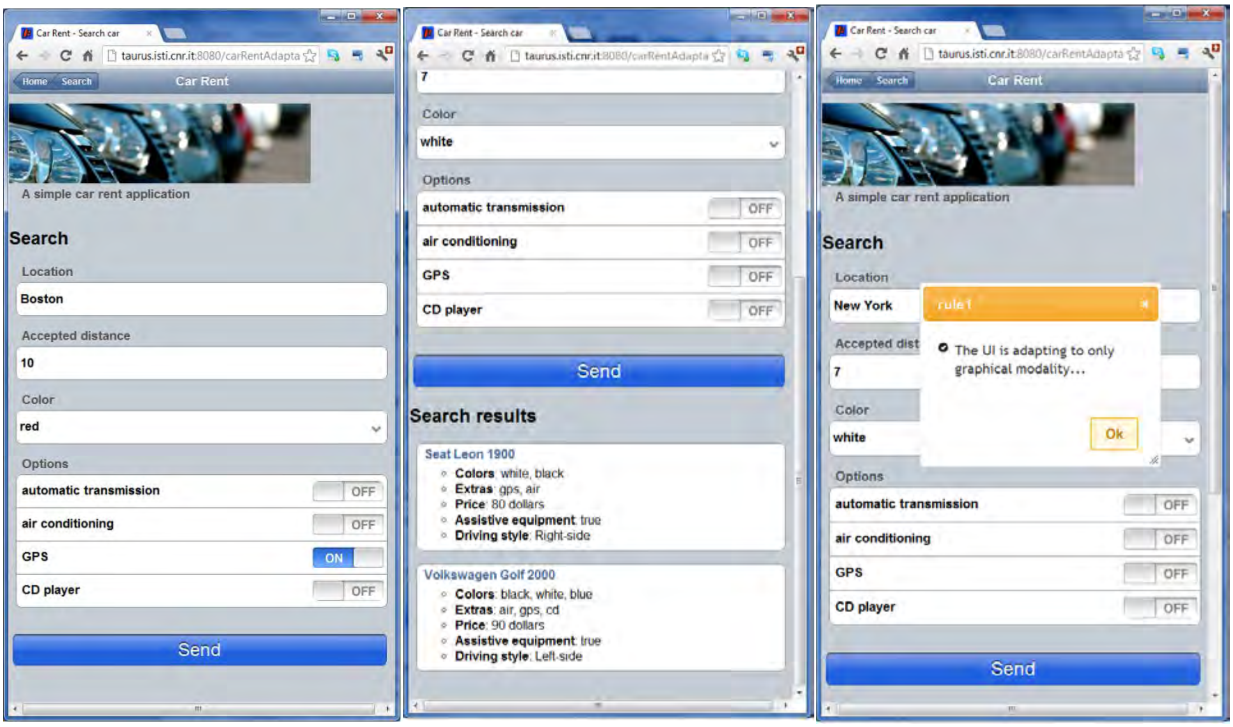
2.3 Conclusions

In the process of AUI development, adaptive rules must be carefully designed and evaluated to avoid usability and user experience pitfalls. Applying UCE in different phases of the development is helpful to detect the flaws of adaptive features in time. On the basis of the results of two user studies, some common drawbacks of adaptive systems are detected and eliminated in our application system. The remedies or potential improvements of some of these drawbacks have been also proposed. As a main result, we came to know that adding user support information can help users to comprehend and accept adaptation rules. Furthermore, we argue that enriching the context and users' profile can increase the precision of adaptation. Also, enabling the user to intervene into the adaptation at any time will improve user experience by improving the controllability. We are convinced that the iterative evaluation of adaptive systems is crucial to the successful development of AUIs. Regarding the iterative testing of such systems, we are happy to report that the fidelity of the testing environment obviously plays no role with respect to the users' rating of the adaptation rules. Thus, rapid iterative testing of adaptation rules does not need to be an expensive enterprise and is therefore highly recommended.

3 CNR Evaluation

A user test has been conducted in order to evaluate the adaptations of a Multimodal User Interface (MUI) that combine graphical and vocal modalities. The multimodal user interfaces evaluated have been obtained through a model-based specification using the MARIA language.

The MUI considered here supports interaction with the Car Rental service, which was used for the integrated demo in the previous review meeting. By using such service the user can perform a search to a Car Rental service with the possibility of specifying various parameters.


Car Rent - Search car
taurus.isti.cnr.it:8080/carRentAdapta
Home
Search
Car Rent
A simple car rent application
Send

a)

Search

Location: Boston

Accepted distance: 10

Color: red

Options

- automatic transmission: OFF
- air conditioning: OFF
- GPS: ON
- CD player: OFF

b)

7

Color: white

Options

- automatic transmission: OFF
- air conditioning: OFF
- GPS: OFF
- CD player: OFF

Search results

Seat Leon 1900

- **Colors**: white, black
- **Extras**: gps, air
- **Price**: 80 dollars
- **Assistive equipment**: true
- **Driving style**: Right-side

Volkswagen Golf 2000

- **Colors**: black, white, blue
- **Extras**: air, gps, cd
- **Price**: 90 dollars
- **Assistive equipment**: true
- **Driving style**: Left-side

c)

Search

Location: New York

Accepted dist: 7

Color: white

rule1

The UI is adapting to only graphical modality...

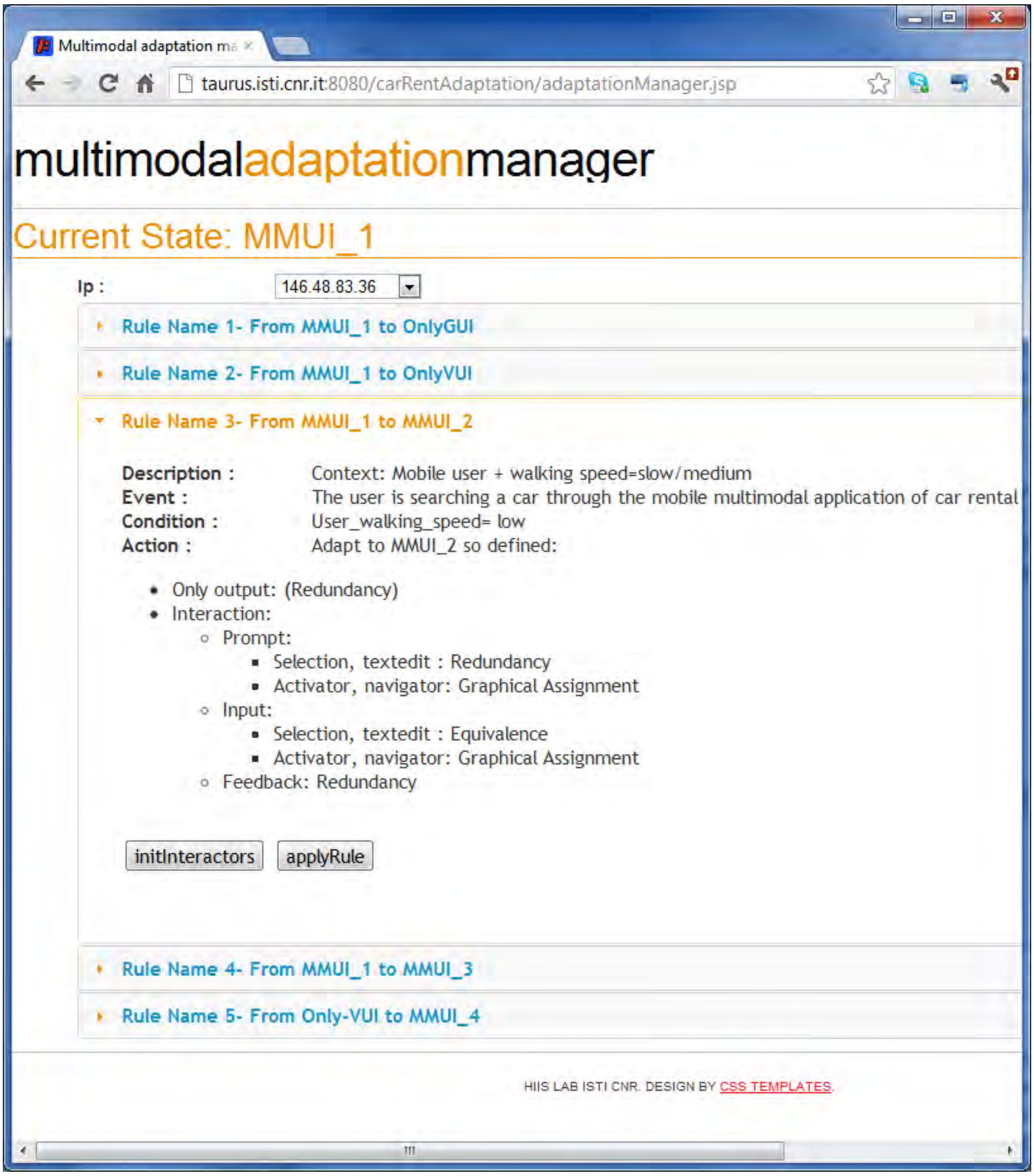
Ok

Options

- automatic transmission: OFF
- air conditioning: OFF
- GPS: OFF
- CD player: OFF

Fig. 7 : a) The UI of the Car Rental application; b)The UI showing some search results; c)The window alerting the user of the adaptation

Figure 7 shows some screenshots of the Car Rental application used for the evaluation. More specifically, the first window (see Figure 7.a) shows the form enabling the user to specify the search. Figure 7.b shows the part of the UI presenting some search results. Figure 7.c shows the window alerting the user that the UI was adapting.



Multimodal adaptation ma
taurus.isti.cnr.it:8080/carRentAdaptation/adaptationManager.jsp

multimodaladaptationmanager

Current State: MMUI_1

Ip : 146.48.83.36

- Rule Name 1- From MMUI_1 to OnlyGUI
- Rule Name 2- From MMUI_1 to OnlyVUI
- Rule Name 3- From MMUI_1 to MMUI_2

Description : Context: Mobile user + walking speed=slow/medium
Event : The user is searching a car through the mobile multimodal application of car rental
Condition : User_walking_speed= low
Action : Adapt to MMUI_2 so defined:

- Only output: (Redundancy)
- Interaction:
  - Prompt:
    - Selection, textedit : Redundancy
    - Activator, navigator: Graphical Assignment
  - Input:
    - Selection, textedit : Equivalence
    - Activator, navigator: Graphical Assignment
  - Feedback: Redundancy

- Rule Name 4- From MMUI_1 to MMUI_3
- Rule Name 5- From Only-VUI to MMUI_4

Fig. 8 :The UI used by the moderator to trigger the adaptation

Figure 8 shows a screenshot of the UI used by the evaluator moderator to trigger the various adaptations considered in the evaluation (in this case, the adaptation rule considered was adaptation 3).

3.1 Participants

Ten volunteers (4 females) aged 20 to 44 (M = 31.2, SD = 6.5) participated in the study. The majority of subjects (7 out of 10) had never used a multimodal (graphical+vocal) user interface before the test. Among the users who already had some experience with multimodal user interfaces, two subjects declared to have already used them in web environments: in particular, one subject specified to have tried multimodal browsers like Opera or NetFront. Another user stated to have tried multimodal UIs in some interactive installations. Subjects were recruited using mainly the professional and social network of the ISTI-CNR team who involved personnel of their institute, as well as some relatives/friends of the institute staff. The educational level of the users was varied: 4 held a PhD, 2 a Master Degree, 3 a Bachelor and 1 a High school degree.

3.2 Adaptation Scenarios and Tasks

Users were asked to try a number of adaptations we identified and applied to the multimodal (graphical+vocal) web user interface of the Car Rental scenario (see Figure 7). More specifically, five tasks were identified. In each of the associated five scenarios, users had to try the Car Rental user interface in two different states: before adaptation and after adaptation. In each scenario, before carrying out the tasks, the user was informed of how to interact with the user interface before adaptation. Then, at a certain moment while executing the tasks, some contextual conditions were supposed as changed (the user was informed about this). As a consequence, an adaptation was triggered. Then the user had to complete the tasks by using the modified user interface (namely: the user interface after adaptation). This was done in order to make each user experiment both the interaction before adaptation and the interaction after adaptation. Most of the times the adaptations changed the way in which the two modalities were allocated to enable user's interaction. For instance, in one case adaptation changed the multimodal (graphical+vocal) UI into an only-vocal UI, while in another case adaptation changed the multimodal UI into an only-graphical UI. Therefore, in both such cases the initial multimodal UI was transformed into a mono-modal UI. Other times the UI remained multimodal before and after adaptation: in these cases the adaptation mainly changed the way in which the two modalities (graphical and vocal) were allocated to user input and/or system output.

Regarding the adaptation rules associated to the various scenarios, we identified them in such a way to cover and analyse different aspects/features of our adaptation approach. In particular, since we consider multimodal UIs, we want to investigate different types of adaptations involving different modalities and then understand whether different allocations of such modalities can result in better UIs in various situations/contexts of use. To this goal we identified three quite 'extreme' cases in which the adaptation totally changes the UI: as you will see adaptation 1 basically transforms a multimodal UI into an only-graphical UI, adaptation 2 changes a multimodal UI into an only-vocal UI. Adaptation5 does the opposite by changing a monomodal, only-vocal UI into a multimodal one. The remaining two cases were aimed to cover other intermediate cases, in which the UI remains multimodal while changing the allocations of the graphical and vocal modalities to the various UI elements. In addition, with adaptation 4 we also wanted to cover the exploitation of information contained in the user model (general preferences and information).

In order to identify how the various modalities are differently allocated, we basically referred to the CARE properties (see (Coutaz et al., 1995)), namely: *Complementarity* (when the considered part of the UI is partly supported by one modality and partly by another one), *Assignment* (the considered part of the UI is supported by one assigned modality), *Redundancy* (the considered part of the UI is supported by both modalities), *Equivalence* (the considered UI part is supported by either one modality or another).

For the test, we selected tasks with low complexity, chosen to mimic activities typically done on this kind of websites (e.g. searches). In particular, users were asked to make a search for a car through the considered Car Rental UI, using the UI before adaptation. Then, when analysing the result, they were supposed to be unsatisfied with it. Therefore, they had to refine the list of results obtained from their search and then modify the search while using the adapted UI.

More in detail, the selected tasks were the following ones:

3.2.1 Adaptation Scenario 1 (Noisy environment)

3.2.1.1 Tasks

1.1 Search for a white car at a maximum distance of 7 km from New York.
1.2 Submit your search and analyse the result. Suppose that you're not satisfied with it.
1.3 Modify the search to get the cars at 10 km from New York, and with GPS on board.
1.4 Submit the new search and analyse the result.

3.2.1.2 Context change

The supposed context change in this scenario is the fact that the environment becomes noisy.

3.2.1.3 Adaptation Performed

In this scenario, at the beginning the user is able to interact with the multimodal UI in such a way that the system output is provided in a redundant manner (namely: both vocally and graphically), while the user provides input only graphically. User performs the first tasks of this scenario (1.1 and 1.2) by using this UI.

Then, we suppose that, before doing the refined search (tasks 1.3 and 1.4), the above context change occurs and thus an adaptation is triggered. As a result of this adaptation, the user carried out tasks 1.3 and 1.4 by using an UI which actually resulted a mono-modal UI: both user input and system output were provided in a graphical manner.

3.2.2 Adaptation Scenario 2 (User walking fast, outdoor)

3.2.2.1 Tasks

2.1 Search for a car at a maximum distance of 10 km from Boston, with GPS on board.
2.2 Submit your search and analyse the result. Suppose that you're not satisfied with it.
2.3 Modify the search to get the cars 10 km far from Boston and having CD Player.
2.4 Submit the new search and analyse the result.

3.2.2.2 Context change

The supposed context change in this scenario is the fact that the user is walking outdoor, quite fast, in a sunny day.

3.2.2.3 Adaptation Performed

In this scenario, at the beginning the multimodal UI is the same as the one initially used in the previous scenario: the system output is provided in a redundant manner, the user provides input only graphically. User performs the first tasks of this scenario (2.1 and 2.2) by using this UI.

Then, when the above context change occurs, the UI is adapted before doing the refined search. As a result of this adaptation, the user carried out tasks 2.3 and 2.4 by using a mono-modal UI: both user input and system output are provided in an only-vocal manner.

3.2.3 Adaptation Scenario 3 (User walking slowly)

3.2.3.1 Tasks

3.1 Search for a blue car 20 km far from Richmond, with air conditioning on board.
3.2 Submit your search and analyse the result. Suppose that you're not satisfied with it.
3.3 Modify your search to only the cars that are at 11 km from Richmond.
3.4 Submit the new search and analyse the result.

3.2.3.2 Context change

The supposed context change in this scenario is the fact that the user is walking slowly.

3.2.3.3 Adaptation Performed

In this scenario, at the beginning the multimodal UI is the same as the one initially used in the previous scenarios: the system output is provided in a redundant manner (vocal and graphical), the user provides input only graphically. User performs the first tasks of this scenario (3.1 and 3.2) by using this UI.

Then the UI is adapted before doing the refined search. As a result of this adaptation, the user carried out tasks 3.3 and 3.4 by using an UI which was still multimodal but having a different allocation of modalities.

Regarding the UI elements just providing output to the user, they were redundantly rendered (both graphically and vocally). Regarding the interaction of the user with the system we distinguished between selection/edit elements and elements supporting navigation/functionality activation.

Selection/edit elements supported user input in an equivalent manner (either vocal or graphical modality), and rendered the feedback given in response to such input in a redundant manner; the prompt (i.e. the UI output representing that the UI is ready to receive an input) was redundantly provided.

For the activation/navigation elements the user could graphically provide input, the prompt was graphically provided as well, and the feedback was provided in a redundant manner.

3.2.4 Adaptation Scenario 4 (Older user)

3.2.4.1 Tasks

4.1 Search for a black car 20 km far from London, with air conditioning and GPS onboard.

4.2 Submit your search and analyse the result. Suppose that you're not satisfied with it.

4.3 Modify your search to get only the cars that are at 9 km from London, with GPS.

4.4 Submit the new search and analyse the result.

3.2.4.2 Context change

With this scenario we consider an adaptation which also exploits the information contained in the user model (e.g. user preferences, user information...). In order to do this, we suppose that the first two tasks are performed by the user by interacting with an UI which does not take into account information about the context/user, while the adapted multimodal UI does.

Such information is that the user is an Italian older man which owns a specific type of car (e.g. a touring car). The tasks foreseen in this scenario imply that the user selects -for his car rental search- a destination in a foreign country having a different driving rule (right-hand traffic country vs. left-hand traffic country).

3.2.4.3 Adaptation Performed

In this scenario, the idea is that in order to facilitate the user in driving in the selected foreign country, the adapted UI will first render the available cars which are more similar to the one owned by the user, exploiting at the same time a different allocation of the graphical and vocal interaction modalities.

Indeed, at the beginning the multimodal UI is the same as the one initially used in the previous scenarios: the system output is provided in a redundant manner (vocal and graphical), and the user graphically provides input.

The adapted UI still enables the user to receive system feedback in a redundant manner, namely the UI elements just providing output to user offer a graphical+vocal information (redundancy). Regarding the interaction of the user with the system, the input is vocally provided, while the prompt and the feedback received in response to user input are redundantly provided.

3.2.5 Adaptation Scenario 5 (User sitting in a bus)

3.2.5.1 Tasks

5.1 Search for a blue car 5 km far from New York. The car should have CD Player onboard.

5.2 Submit your search and analyse the result. Suppose that you're not satisfied with it.

5.3 Modify your search to get only the cars that are at 5 km from Cambridge, with GPS.

5.4 Submit the new search and analyse the result.

3.2.5.2 Context change

The supposed context change in this scenario is that the user is sit in a quite noisy environment (e.g. he finally sit down in a bus after having walked quite fast to reach it).

3.2.5.3 Adaptation Performed

In this scenario, at the beginning the UI is a mono-modal one: the user input is provided only in a vocal manner, and the same holds for the system output. User performs the first tasks of this scenario (5.1 and 5.2) by using this UI.

In the adapted UI used for tasks 5.3 and 5.4, the system output is provided in a graphical manner. For the user interaction we had that the user input and the associated prompt was provided graphically, while the feedback produced by the system in response to user input was provided in a redundant manner (both graphically and vocally).

3.3 Methodology and Procedure

The study was a within-subject one (each user had to analyse all the adaptations). In order to effectively experiment all the adaptations, in each adaptation scenario each user performed twice the search task. In order to diminish carryover effects, half of the users experimented the adaptations starting from the first rule, while the remaining half started from the last one. A moderator was also available during each user session.

The test was carried out by using the Google Chrome browser on a laptop through which it was possible to connect to the desktop PC hosting the Car Rental application. Since we had limited time available and the prototype was not yet well-engineered the test was done in laboratory. In order to simulate an interaction through a mobile device, the size of the browser window exploited by the user was appropriately resized.

The evaluation was composed of various parts. The study was initiated with a short presentation in which general information and instructions were given to each subject by the experiment moderator. To this goal, the user was provided with a brief introductory test to read. In this text there was a description of the main ideas of the multimodal adaptation approach, as well as an overall picture of the tasks they were expected to perform, the scenario domain, and the kind of adaptations they were going to experiment. After reading this text, a demographic questionnaire was applied: each participant answered to a set of general questions about their education, age, and experience in using multimodal user interfaces.

The majority of users were novices with using multimodal user interfaces. Thus, in order to familiarise with them, after reading the introduction each subject performed a set of practice tasks equivalent to those that would be presented to them in the main part of the experiment (fill in a form contained in an application different from the one used in the test and then submit it), and using a multimodal (graphical and vocal) user interface.

After this, each participant started the test. For each of the five adaptation scenarios identified, participants had to read the associated description explaining how to interact with multimodal UI to carry out the described tasks in the context supposed (simulated) in each adaptation scenario. Each adaptation scenario basically consisted of performing two search tasks: one had to be carried out using the UI before adaptation was triggered, the second one had to be performed after adaptation. It is worth pointing out that the adaptation was actually manually triggered by the moderator. After experiencing the adaptation users had to rank the type of adaptation just experienced by filling in a specific part of the questionnaire devoted to assess the adaptation according to a number of criteria. They also had the opportunity to give further comment on the adaptation just tried. After filling in the questions about the last adaptation, they were asked to provide some further suggestion on the overall approach (strong points, weak points).

We collected user's feedback on the various adaptations that were proposed. In particular, for each adaptation scenario we asked to rate the underneath adaptation rule according to a number of evaluation criteria that identified in previous Serenoa deliverables (see D2.4.2 and D2.4.1). It is worth pointing out that just a subset of the criteria reported in those deliverables was selected for the experiment. This selection was done by also taking into account the conditions in which the experiment was carried out, especially the fact that some conditions were just simulated.

Finally, the evaluation criteria considered were:

- *User's awareness of the* adaptation: the goal is to understand to what extent the user was able to realise that a change in the UI was caused by adaptation;
- *Appropriateness of the adaptation selected*: when the adaptation decision comes up, it can be applied in different ways (e.g., different colours, different layouts, different modalities,..). The user was asked whether the system selected a good/appropriate adaptation strategy;
- *Continuity of the adaptation* : the user was asked to rate to what extent it was easy to continue the

interaction after adaptation;

- *Transition of the adaptation*: it refers to the behaviour of the user interface during the adaptation. The goal was to assess to what extent the adaptation process was understandable and allowed users to realise what is happening.
- *Impact of the adaptation in decreasing the interaction complexity*: one possible advantage of the adaptation is to get a decrease in the interaction complexity of the system;
- *Impact of the adaptation to increase the user satisfaction*: the goal was to understand to what extent, the adaptation just experienced increased user's satisfaction in exploiting the UI.

Each criterion had to be ranked according to a 1-5 scale (1 is the worst rating, 5 is the best one), therefore we computed the Mean ± SD (Standard Deviation) of ranks for each rule and for each criterion, in order to understand which rule was better received by users.

3.4 Results and Discussion

In this section we summarise the results gathered by analysing the data collected during the test, mainly from the filled questionnaires. In addition, we also report on additional feedback/remarks/suggestions provided by users.

All the users successfully completed the expected tasks. Regarding the time needed to carry out the tasks, it was recorded for each scenario by the moderator. Figure 9 shows the results by visualising the seconds needed by users, on average, to complete the tasks associated to each adaptation scenario. As you can see, adaptation 5 was the adaptation that on average needed more time (M=155, SD=67), while Adaptation 3 was the best one in terms of time needed (M=147, SD= 50). This may be explained by the fact that adaptation 1 and 2 switched the interface from a multimodal to a single modality (respectively graphical and vocal) and assumed quite simple adaptation scenarios. Therefore it was easy to associate the context cause to the adaptation effect, also for the users that started the test from adaptation 1. In addition, the users that started from the adaptation 5 were significantly quicker in completing the adaptation1 and 2 that the ones that started from adaptation1, and this leads to a mean value close to the best users' performance in task 3. Considering that for both groups the adaptation 3 was always in the middle of the test, even if the adaptation changes were more difficult to be perceived with respect to the two first tasks, all the user were used to the interface using both modalities, and this explains their performance. In adaptation scenarios 4 and 5 the effects of the adaptation were more difficult to be grasped, since the simulated situation was not trivial. In particular, the adaptation scenario 5 started from the only-vocal modality, and it was more difficult for people who had never seen the graphical rendering of the car search form to figure out what the application was requesting them to provide. This mainly explains why (as we expected) adaptation5 was the one with the worst users' performance.

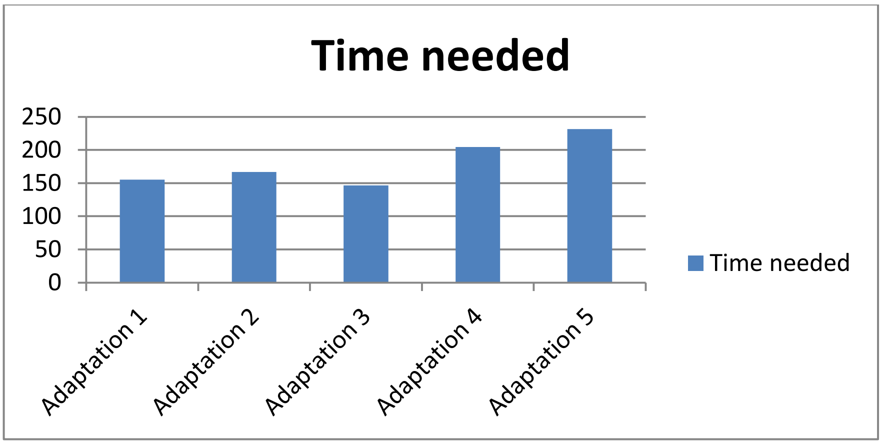


Figure 9: Time needed for adaptation rules

In the questionnaires, a 1-to-5 semantic differential scale was used by the participants to provide their ratings on various aspects of adaptation. In that scale, 5 was the most positive score, and 1 was the most negative one. So, the meaning of the scores was the following: 1 = very bad, 2 = bad, 3 = neutral/ambivalent, 4 = good, 5 = very good.

User's awareness of adaptation – The adaptation which received the best score was Adaptation 2 (M = 4.9, SD = 0.3), while the one which received the worst score was Adaptation 1 (M = 4.1, SD = 0.7).

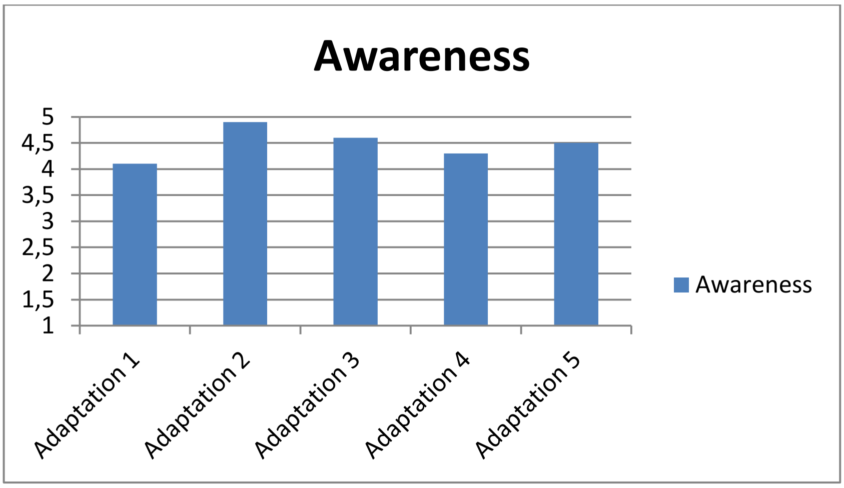


Figure 10: User's awareness of adaptation rules

The overall user awareness of the adaptation in the considered adaptation scenarios is very good, all of them are above 4. In particular, it is possible to notice that for adaptation 2 the awareness is very high. This may be explained with the fact that the transition in this task was from the graphical + vocal modality to the only vocal one. Therefore, for the users, was clear that something changed in the user interface, since they were not able to see anything on the screen. In particular, one user was a bit puzzled by the fact that the graphical UI disappeared, since s/he did not read carefully the transition message while another one, after starting the interaction asked whether there was a possibility to refuse the adaptation and go back to the graphical modality.

Appropriateness of adaptation – The adaptation which received the best score was Adaptation 1 (M = 4.5, SD =0.7), while the one which received the worst score was Adaptation 4 (M = 3.6, SD = 1.2).

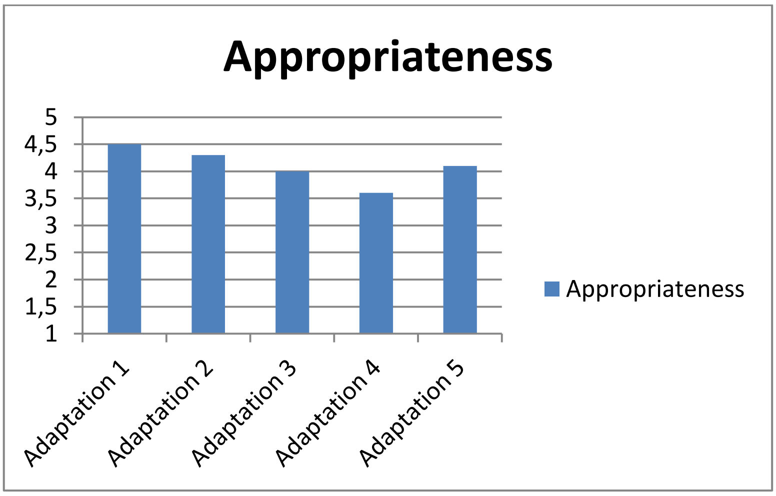


Figure 11: Appropriateness of adaptation rules

The overall assessment of the appropriateness of the adaptation is good. Apart from the analysis of the numerical rating of this adaptation aspect, in this case the free comment field at the end of each task provide us with some useful hints for the management of the different modalities when adapting the UI. In general, the users considered appropriate the fact that the UI may switch to the complete vocal modality, given the contextual situation supposed. Nevertheless, considering a search task like the one proposed in the testing scenarios, it is important to provide the user with the opportunity of comparing different results and the possibility of entering inputs and/or randomly explore the search list. Such kind of information browsing is difficult to support through the vocal modality. Therefore some users may prefer to maintain the modality that suites better to task they want to complete rather than the one that suits better to the context. Such point affected in particular the adaptation 3, where the graphical UI is enhanced with the possibility to provide vocal feedback and inputs because the user is walking slowly. For this adaptation, one user commented that "*Even if I could be walking slowly, I may not be interested in receiving voice feedback. This may have to do with personal preferences. Some users may be more familiar with current modalities and find them friendlier*". This is an indicator that the change of modality that is not familiar to the user may represent a higher load for the user if compared to a simultaneous handling of a well-known interaction modality plus some external activity.

If we consider the adaptation 4, the user is supposed to be an older person and, after the adaptation, it was possible to graphically select which field to complete, but the input should be provided vocally. Even if the user were aware of how to interact with the interface, such solution was not immediate to understand: the moderator observed that after clicking the input field, the users started typing the value. Instead, if the fields were read from the text to speech automatically (the page in the vocal modality is read in a continuous way), the user replied vocally to the prompt. The mixture of modalities during the execution of a simple task like completing a field (graphical field selection plus vocal completion) has been evaluated as less appropriate than the other proposed adaptations.

Continuity of adaptation – The adaptation which received the best score was Adaptation 1 (M = 4.9, SD = 0.3), while the one which received the worst score was Adaptation 4 (M = 4, SD = 0.9).

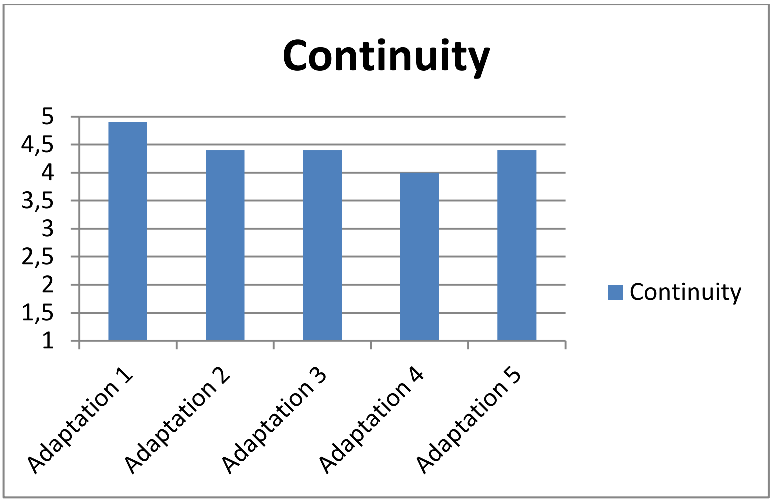


Figure 12: Appropriateness of adaptation rules

The continuity of adaptation was rated very well for all the tasks. It is possible to comment on the maximum and on the minimum value, in order to understand the distance from the other ones. With respect to the continuity supported in adaptation 1, it is possible to say that since the adaptation turned off the redundancy of the output interactors (which initially were both visible on the screen and read by the text to speech), all the user started entering the input graphically and continued smoothly after the adaptation, even if they noticed that the application did not provide vocal feedback anymore.

The minimum value is again registered for adaptation 4, and it is a consequence of the discussion on the adaptation appropriateness. Mixing the modalities for selecting and completing a field has been perceived as a factor that in some way interrupted the way the users interact with the application before and after the adaptation.

Transition of adaptation – Three adaptations received the best score: Adaptation 1, Adaptation 2 and Adaptation 3 (M = 4.6, SD = 0.5), while the other two received the worst score: Adaptation 4 (M = 4.4, SD = 0.5); Adaptation 5 (M=4.4, SD=1).

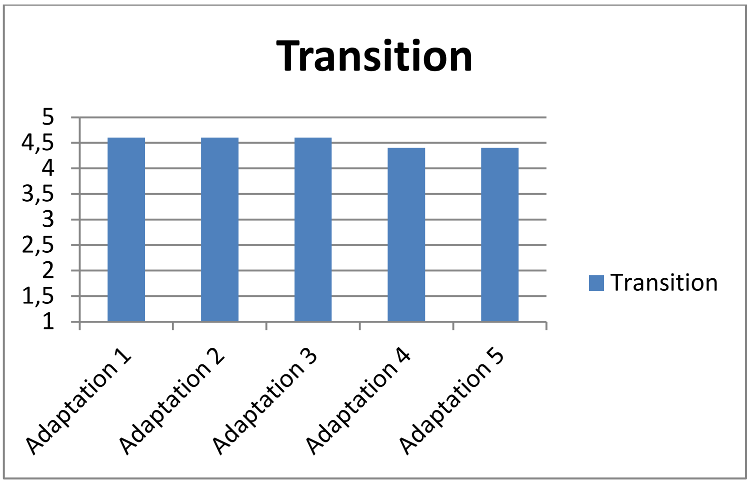


Figure 13: Transition of adaptations

The transition between the initial and the adapted version of the user interface was rated very good. The users appreciated the fact that the application showed a message before changing its behaviour and the modality. When switching between modalities, and in particular from graphical to desktop, some users found natural to receive a graphical message, while some others complained that, if they were supposed to be walking fast (task 2), they should have received a vocal notification. In such cases, it is probably better to have a redundant message in both modalities.

Impact of adaptation in decreasing the interaction complexity – The adaptation which received the best score was Adaptation 5 (M = 4.6, SD = 0.7), while the ones which received the worst score were Adaptation 1 (M = 3.7, SD = 1.2), Adaptation 3 (M = 3.7, SD = 0.8), and Adaptation 4 (M = 3.7, SD = 1.3).

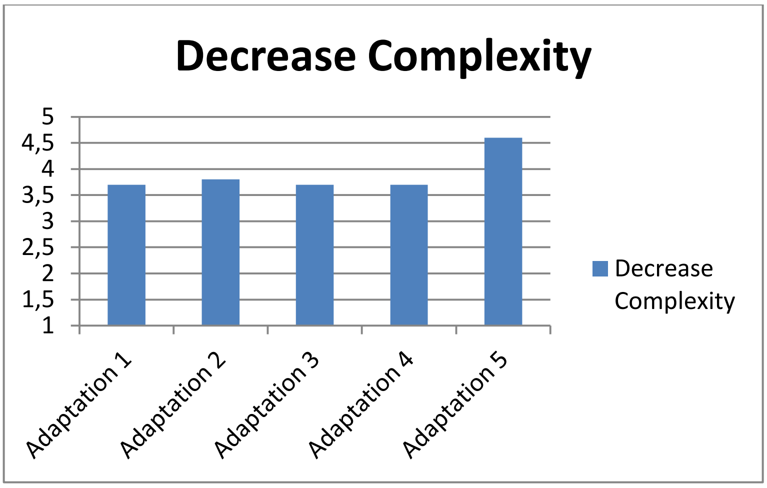


Figure 14: Impact of adaptations in decreasing interaction complexity

Impact of adaptation in increasing user satisfaction – The adaptation which received the best score was Adaptation 5 (M = 4.4, SD = 0.7), while the one which received the worst score was Adaptation 3 (M = 3.8, SD = 0.9).

The users found that there was a more than sufficient decrease in complexity after the adaptation. The higher value was related to the adaptation 5, where the user started from an only-vocal version of the UI and the interface changed to a graphical version + vocal feedback. The reason of this may be explained by the user's unfamiliarity with the vocal interaction.

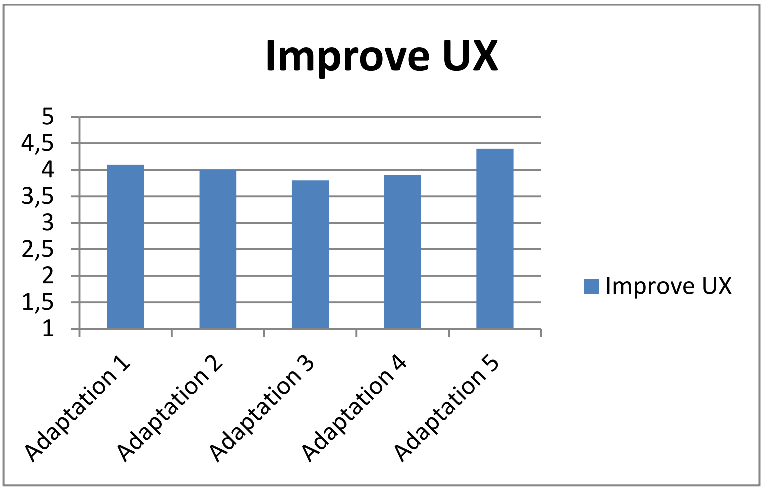


Figure 15: Impact of adaptations in improving user experience

Overall rating – An overall rating was calculated by taking into account the ratings that were received by each adaptation rule on the various aspects presented beforehand (e.g. appropriateness, awareness, ..). From this calculation it was derived that the adaptation which received the best score was Adaptation 5 (M = 4.4, SD = 0.7), while the one which received the worst score was Adaptation 3 (M = 3.8, SD = 0.9).

Further comments on each adaptation rule

- *Adaptation 1*

Regarding adaptation 1, a user noticed that an even better adaptation could have been reached by reducing the amount of available information presented in the resulting graphical UI, by just putting only the most relevant information available. This would also depend on whether the user is moving or stopped: in a fixed position, no changes to the amount of information presented should have been required. In noisy and mobile environments, the attention is reduced and the interaction should focus on the most important things. The same user found this adaptation rule the most convincing case of adaptation after a change of context.

- *Adaptation 2*

Regarding this adaptation, four users suggested that in the supposed scenario, a graphical rendering of the search result would have also been useful to guide the users and quickly provide them at a short glance information about the state of their current operation. Another user suggested adding a sort of “Replay” command in order to better understand something that was already said. Another user found this adaptation as fully justified by the supposed context

- *Adaptation 3*

Regarding this adaptation, one user noticed said that, although it was possible to select elements also through vocal input, he personally prefers to use the graphical UI because it's faster than voice. Another user noticed that further context variables could be considered on whether or not to have voice feedback. He noticed that even if he could be walking slowly, he may not be interested in receiving voice feedback, and therefore this may have to do with personal preferences as some users may be more familiar with current modalities and find them friendlier. Another said not to have any further remark as he considered this adaptation as the bet one in the test.

- *Adaptation 4*

Regarding this adaptation, one user noticed that accepting only vocal input in a standing scenario may be frustrating if recognition is not working well. Therefore, he suggested that in this case it could have been good to leave both options available - graphical and vocal. Another user said that he found really annoying the adapted UI in which the input could be provided only through a vocal modality. Another user also suggested to have input both vocally and graphically. Another user noticed that an old Italian user would probably not receive positive results from a system using only inputs with an English accent.

- *Adaptation 5*

Regarding this adaptation, one user complained on the fact that when he submitted the form after adaptation (by using a graphical modality), he did not receive any feedback. Actually, in the proposed prototype, the results were always written in the same part of the UI and therefore there was the possibility that some results just overwrote the list of the results shown in the first search, therefore the change went unnoticed.

Further Overall Comments

One user said that the vocal interaction should be improved in order to allow the user to have a faster and more "natural" interaction. He noticed that vocal modality still needs improvements although the recognition is getting better, and also mentioned the possibility to create new modes to navigate with vocal input to speed up the process by possibly considering erase or undo options. We also received suggestions for adding a “replay” command. Another user said that "vocal can be useful if the interaction required really simple tasks due to the reduced vocal memory of human beings. It requires an additional effort during the interaction.” Therefore he suggested considering also gestures and touching modalities. It was also suggested by a different user that the list of the search results should be always graphically rendered since the final choice may be among several cars and as a car renter, he would probably need to compare different cars and prices.

One user appreciated the (simulated) context-driven adaptation. Another user noticed that "if a person does not notice the message alerting the incoming adaptation, then it could be difficult to continue the proper interaction”. One user was just convinced of adaptation scenario 1, another one most appreciated adaptation 3. Finally, a user find it not hard to interact with the vocal modality but she declared to still prefer the graphical one, while another one appreciated the possibility to choose the preferred interaction modality.

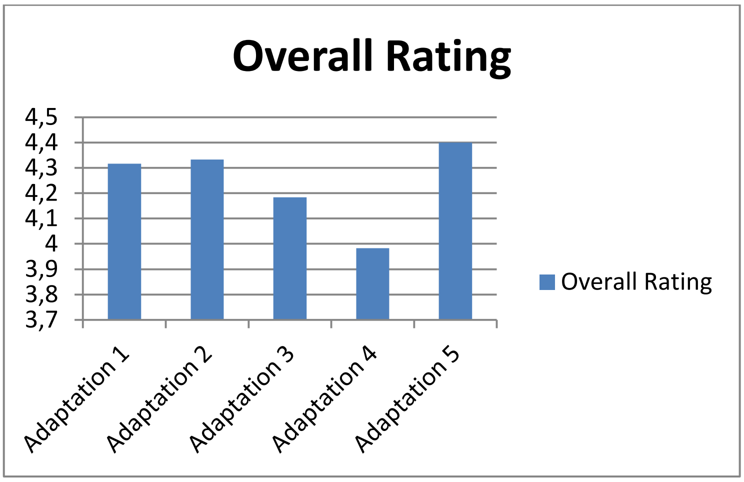


Figure 16: Overall rating

Overall, the results of this test suggest that users appreciated the adaptations offered in the various scenarios we identified. This was very positive, also taking into account that the majority of them had never used a multimodal (graphical and vocal) user interface before the test.

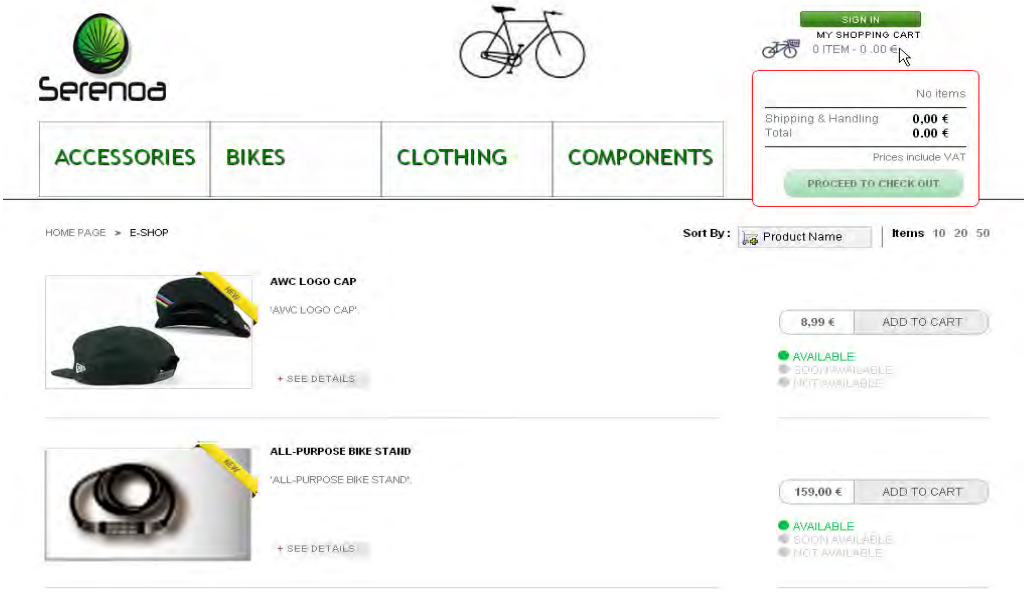
4 W4 Evaluation

The W4 prototypes aim at using Serenoa's adaptation rules on 'classical' business applications: one Business to Customer (B2C) prototype (wide public, customer oriented, without prior knowledge about the application), one Business to Business (B2B) prototype (typically, employees of an SME with good knowledge and frequent usage of the application).

The E-commerce scenario is based on the idea of a bicycle online shop selling bikes and bike related equipment or parts.

The default application (prior to adaptation) presents:

- A web application accessible on the Internet, for the customers (sees products, descriptions, price, add product to a basket and order the products). Additionally, a page will show the order status but this status is updated by the other modules.



AWC LOGO CAP

'AWC LOGO CAP'

+ SEE DETAILS

8,99 € ADD TO CART

● AVAILABLE
● SOON AVAILABLE
● NOT AVAILABLE

ALL-PURPOSE BIKE STAND

'ALL-PURPOSE BIKE STAND'.

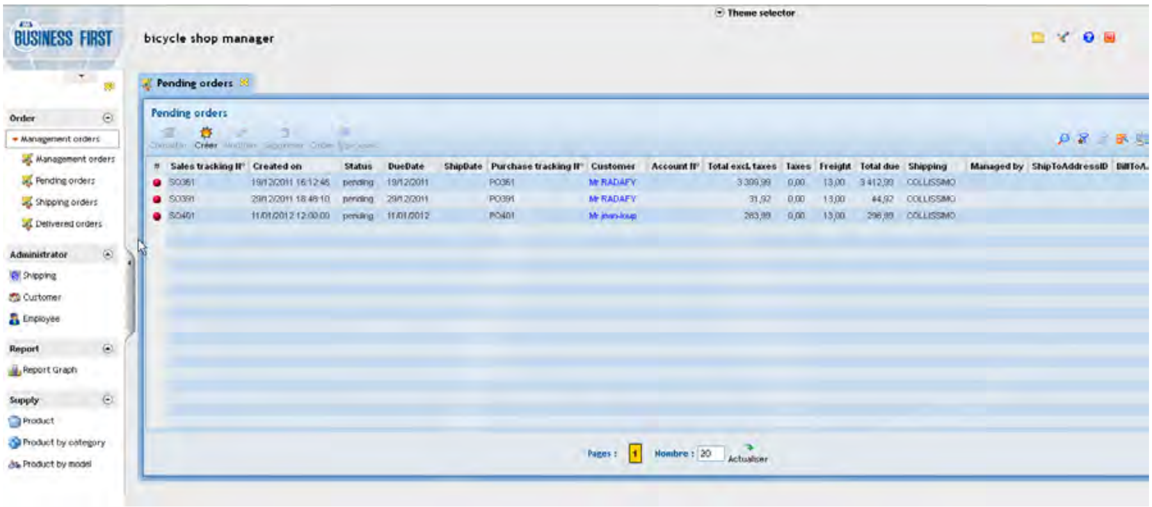
+ SEE DETAILS

159,00 € ADD TO CART

● AVAILABLE
● SOON AVAILABLE
● NOT AVAILABLE

Figure 17: The application considered in the e-commerce scenario

- A web application accessible from the intranet of our bicycle shop, sharing the same database and content, in order to validate the customer order, checks product availability and request for shipping.



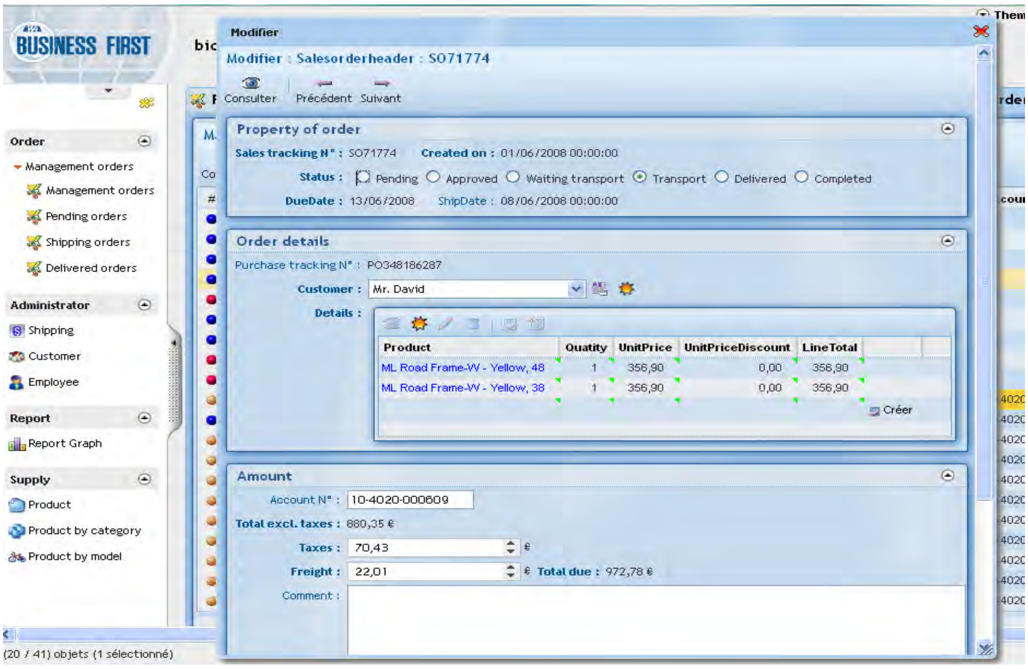
Pending orders

#	Sales tracking N°	Created on	Status	DueDate	ShipDate	Purchase tracking N°	Customer	Account N°	Total excl. taxes	Taxes	Freight	Total due	Shipping	Managed by	ShipToAddressID	BillToA..
●	SO361	19/12/2011 16:12:46	pending	19/12/2011		PO361	Mr RADAFY		3 399,99	0,00	13,00	3 412,99	COLLISSIMO			
●	SO391	29/12/2011 18:48:10	pending	29/12/2011		PO391	Mr RADAFY		31,92	0,00	13,00	44,92	COLLISSIMO			
●	SO401	11/01/2012 12:00:00	pending	11/01/2012		PO401	Mr jean-loup		283,99	0,00	13,00	296,99	COLLISSIMO			

Pages : 1 Nombre : 20 Actualiser

Figure 18: UI for checking product availability

- A small automated process will simulate the shipping process and change the order status (prepares, shipped, received ...)



Modifier : Salesorderheader : SO71774

Property of order

Sales tracking N° : SO71774 Created on : 01/06/2008 00:00:00

Status : ○ Pending ○ Approved ○ Waiting transport ● Transport ○ Delivered ○ Completed

DueDate : 13/06/2008 ShipDate : 08/06/2008 00:00:00

Order details

Purchase tracking N° : PO348186287

Customer : Mr. David

Details :

Product	Quatity	UnitPrice	UnitPriceDiscount	LineTotal
ML Road Frame-W - Yellow, 48	1	356,90	0,00	356,90
ML Road Frame-W - Yellow, 38	1	356,90	0,00	356,90

Amount

Account N° : 10-4020-000609

Total excl. taxes : 880,35 €

Taxes : 70,43 €

Freight : 22,01 € Total due : 972,78 €

Comment :

(20 / 41) objets (1 sélectionné)

Figure 19: The UI for checking the status of the order

In this first report we propose to evaluate the suitability of several adaptation rules to these prototypes.

4.1 Methodology

4.1.1 Participants

First participants are students having an internship within W4 company, but the evaluation is still recruiting additional external participants without any prior knowledge of our technology.

4.1.2 Evaluation criteria

The evaluation criteria are the same as the TID's prototype described in section 5.1.2 (Evaluation criteria).

The evaluation form is also shared in order to have more consistent results between partners. We use the form described in Annex A.1 Questionnaire.

4.1.3 Procedure

The E-Commerce scenario will be installed on an external website (accessible to participants from any location). The public website bicycle shop prototype will be accessible without any prior login; authentication will be only requested when the participant purchases a product. He (she) will be able to create an account online.

The corporate web application (B2C: Bicycle shop manager application) will also be accessible from internet (though a real production application which should be limited to an intranet access) in order to make the evaluation easier for participants. However, an authentication (login / password) for the roles described within the evaluation process below will be given when starting the evaluation.

"Mobile" devices like Android tablet/phone or iPhone/iPad will be accessible from any compatible device: URLs and authentication procedure will be provided to participants (Using Wi-Fi or eventually 3G connexion).

The evaluation process is divided in 5 steps. After each step, a common questionnaire (described below) is proposed to the evaluator about this adaptation and his/her overall impressions and usage experience. Each step will be evaluated individually. When the evaluation form is complete, the evaluator proceeds to next evaluation step.

Thus, the procedure is planned as follows:

- Facilitator introduces the main goal of the study.
- Description of the role of the participant and information to start the test will be given (URL, authentication, installation procedure).
- After each step, the facilitator stops for a moment and presents to user a form which collects user's opinion about the evaluation criteria listed in Section 5.1.2. Once user has filled out this questionnaire, facilitator goes on with the story till the next adaption rule where the process is repeated.

4.2 Adaptation rules

4.2.1 Bicycle shop public web site

Erik is from The Netherlands, living in an Amsterdam neighbourhood. Like most his fellow countrymen, Erik is quite fond of bicycle. His mother language is Dutch but Erik is rather fluent using English and usually browses on-line shops to find good (and cheap) products for his week-end leisure activity: cycling.

He has just discovered the bicycle shop website and browses the on-line product catalogue at home (using a web-browser on a laptop).

Unsure whether to buy the product or not, Erik has to leave home and must to go to work, without ordering the product. In the bus, Erik thinks about a product he has seen on the website. Erik connects to the bicycle shop website with his mobile phone (a quite recent smartphone with a web browser and 3G data access). He navigates to find the product description and additional details. He wants to find the exact colour of the product from the textual description (Erik is colour-blind: the image is not completely sufficient). While navigating, Erik sees a link to a colour-blind adaptation of the website, and of course, interested by this option, tries to activate it.

At the end of the day, he decides to purchase the product from home: Erik selects the product, activates the colour-blind mode, adds the item to the basket, fills in the purchase order form, edits his own credit card information and finally validates his purchase order.

- Mobile web adaptation
    - Event: Public website is access from a mobile device.
    - Condition: web browser is recognized as a mobile application.
    - Action: runtime adaptation of the page content to improve usability and user experience.

- Colour-blind adaptation
    - Event: Public website option is toggled on by user choice.
    - Condition: the user is colour blinded and activates this mode.
    - Action: Images and colours are adapted in order to improve user experience

	Event	Condition	Action
Mobile web adaptation	Public website is access from a mobile device.	Web browser is recognized as a mobile application	Runtime adaptation of the page content to improve usability and user experience.
Colour-blind adaptation	Website 'colour blinded' option is toggled on by user choice or profile	The user is colour blind and activates this mode.	Images and text and page colours are adapted in order to improve user experience

Table 2: Specifying the ECA rules (Mobile web adaptation and Colour-blind adaptation)

4.2.2 Bicycle shop manager: order validation

Juliette is working for the bicycle shop company. She is French but works in Luxembourg. She is in charge of the validation process of customer orders. She has just received the purchase order from Erik. She opens the order form and check if everything is fine before sending the order form to the shipping team. She connects to the corporate website, selects French (her mother language) displays the order list and validates the order. Erik receives a validation email to confirm that his order has been approved.

	Event	Condition	Action
Language adaptation	a multilingual UI has been accessed.	User's preferred language is French	The system displays the information in French

Table 3: Specifying the ECA rules (Language adaptation)

4.2.3 Bicycle shop manager: customer representative

Peter is the customer representative; he is German but currently travelling to the company's headquarters in Luxembourg. He would like to see the latest orders and sales figures. He connects to the corporate application with his Android tablet, using the mobile android adaptation. After authentication, Peter is enabled to visualize daily orders and details, and some graphical charts about monthly sales.

	Event	Condition	Action
Mobile tablet adaptation (Android)	The corporate application is accessed through an android tablet	the Android application is used	The system uses native Android screens to display information

Table 4: Specifying the ECA rules (Mobile tablet adaptation - Android)

A second optional test (if available or depending on 'tablet' availability may be performed using an IPhone or IPAD): same results but using IOS instead of android.

	Event	Condition	Action
Mobile tablet adaptation (IOS)	The corporate application is accessed through an apple mobile device	the IOS application is used	The system uses native IOS screens to display information

Table 5. Specifying the ECA rules (Mobile tablet adaptation - IOS)

5 TID Evaluation

Telefónica I+D is working on the deployment of two prototypes which take advantage of the automatic adaptation offered by Serenoa framework. These prototypes are aligned with both pilot projects developed by Telefónica I+D, namely: HealthDrive and SARA project.

In this first report we propose evaluate the suitability of several adaptation strategies to each prototype.

5.1 Methodology

5.1.1 Participants

The target groups of costumers for HealthDrive and SARA services are quite different. The former are, in principle, anyone who is interested to manage their medical information, although in this first version of the prototype the service is focused in pregnant women. The latter are chronic patients who should monitor their health status daily.

To this first evaluation we have selected two groups of five people each one. One of them is made up of five women between 30-40 years who are or had been pregnant. This group will be part of the HealthDrive evaluation. The second group is made up of five people recruited from the TID's office, three men and two women. They have no chronic illness.

We would like to point out that although users profiles are not totally aligned with the target market, we expect to receive insightful feedback to guide the implementation work from now on. The final evaluation, due to the end of the project, is planned to include a fully functional prototype and users more fitted to target market.

5.1.2 Evaluation Criteria

In the document *"D2.4.2-Criteria for evaluation of CAA of SFEs (R2)"* a comprehensive list of evaluation factors, which are relevant to evaluate the quality of context-aware adaptation of SFEs, is presented. In the current evaluation we deal with those user-oriented aspects closely related with the adaptation strategies followed in TID's case study.

Next we present our proposal of factors which have been taken into account into this evaluation (see *"D2.4.2-Criteria for evaluation of CAA of SFEs (R2)"* for further details about the selection of these factors):

- *Appropriateness:* the aim of this criterion is to understand whether the system selected a good/appropriate adaptation strategy. Briefly, it measures how the adaptation matches the mental model of the user. For example, if users would consider that quality of the interaction has improved when avatar is showed with a sequence of images if device changes.
- *Timeliness of the adaptation:* it refers to the application of adaptation in a timely manner (e.g., not too late, not too early) when there is a need to change some aspect of the user interface to better support the user. For example, if avatar's behaviour is modified in the right moment, that is, when interaction errors are detected or it would be better to keep avatar's personality.
- *Continuity:* it means the possibility to easily continue the interaction after adaptation. At this stage we are interested on the users' perception of continuously interacting with the system. For example, how is perceived the adaptation of the avatar engine when users change their devices. Do they consider the session has not been lost?
- *End-user disruption caused by adaptation:* This criterion evaluates the amount of end-user disruption/frustration caused by the adaptive behaviour. We want to know if pro-activity to switch the modality from voice to only text is annoying.
- *Impact of adaptation on user experience:* to understand to what extent adaptive behaviour (in terms of e.g. adaptation rules) can be effective with regard to user experience. For example, in case of a user has visual difficulties, the activation of avatar's voice would improve the user experience.
- *Impact of adaptation on user performance:* to what extent the adaptation is able to decrease the interaction complexity and then have some positive effect on the user's performance. For example, if

the avatar is considered useful for guiding the interaction and don't represent a higher complexity.

- *Impact of adaptation on error-prevention:* the necessity of making user interactions less prone to errors that can have some critical effect. Thus, in this case the goal of adaptation could be to avoid (or at least reduce) the risk that the user can face critical situations. For example, an error prevention strategy via adaptation would be to implement a more empathic attitude in the avatar if some interaction errors occur.
- *User's perceived confidence and trust in the adaptation:* this factor is about the user confidence in the ability of the adaptive system to predict future needs. It is related with user's concerns regarding privacy, user control, consistency, and system competence. For instance users could be worried if some sporadic environmental noises could trigger the change of modality when they really don't want it.
- *Consistency of the across-device adaptation:* this criterion refers to the level of consistency between the UI design before an adaptation and after an adaptation to a different device. Avatar adaptation from desktop to mobile device is an example of an across-device adaptation where UI harmony needs to be maintained.
- *General likeability:* it is the intention of a person to use a particular system. This criterion seems to have relationship with the perceived usefulness of the system, the easiness in using the system itself and also the likeability of the system (to what extent the user liked/was satisfied by using the system) that the user might have perceived the first time s/he interacted with the system. Although it is difficult to assess this aspect taking into account that some participants are not part of the target group, we encouraged them, as pointed out before, to play a role aligned with the evaluation's goal.

To collect user feedback about all these criteria we have prepared a web form[1] which is passed to participants as next section indicates.

5.1.3 Procedure

The goal of this evaluation is to assess the quality of the adaptation rules as subjectively perceived by the participants. To do that, functional prototypes as well as visual material (i.e. slides and mock-ups) have been prepared to get involved participants in the e-Health scenarios (i.e. HealthDrive and SARA). Notice that these prototypes are not yet fully functional, so in some cases capacities have been simulated or have been showed through support material (see Annex A.2).

Thus, the procedure is as follows: facilitator introduces the main goal of the study and then, a laptop is presented to participant. In this computer is running a very first release of the HealthDrive prototype. Facilitator describes the aim of the application and then starts to tell the scenario presented in section 5.2.1. When an adaptation situation is reached (e.g. language adaptation) then facilitator stops for a moment and presents to user a form which collects user's opinion about the evaluation criteria listed in section 5.1.2. Once user has filled out this questionnaire, facilitator goes on with the story till the next adaption rule where the process is repeated.

The scale of the form has been designed as a Likert scale of 5 levels.

5.2 Adaptation rules

5.2.1 HealthDrive

HealthDrive is a pilot program aims to leverage on consumer devices such as computers, tablet PCs and phones to provide its users access to their personal file on the Andalusian health system. In order to do so, all medical information is digitized and shared by the institutions, with a publicly accessible interface for each user in which she/he can interact with doctors and see their health records. The Andalusian Health System is the official public health system for the Andalusian region in Spain, providing universal health care to its nearly 8.5 million inhabitants. The system is currently in the process of having its centres and processes completely digitalised to provide a faster and more efficient service to its beneficiaries. Telefónica I+D is one of the major entities providing expertise to the public office in order to advance towards its objectives.

[1] https://docs.google.com/spreadsheet/viewform?formkey=dHdjMWR2eVlzaDJqU2ltT2JUbmhFWHc6MQ

Next a use case is described in order to better introduce the application of the adaptation strategies which are going to be part of the study.

*Jane is American and she is living at Granada. She is pregnant and she is using HealthDrive service to be informed about the progress of her pregnancy. HealthDrive let her to access this personal information in her own language (**Language adaptation**). She usually consults her HealthDrive desktop application at home. When she is outside and she tries to access to the eHealth assistant using her smartphone, the avatar presentation is degraded and a sequence-of-images version is presented (**Avatar adaptation**). Afterwards, on the bus, a high level of noise is detected and the avatar voice is no longer audible. The avatar suggests changing the modality and Jane agrees (**Noisy environment adaptation**). Besides she doesn't want people on the bus were aware of the interaction. Then the volume is turned off.*

In Table 6 we follow ECA (Event-Condition-Action) format to specify the context and subsequent actions which are associated to each adaptation strategy.

	Event	Condition	Action
Language adaptation	A multilingual UI has been accessed	The user is American	The system shows and plays the information in English
Avatar adaptation	The user device has changed	The new device video capabilities don't let avatar engine work properly	The avatar is displayed using a sequence of representative images
Noisy environment adaptation	The environment gets noisy	The noise level gets higher of a certain threshold	The application turns off the voice modality

Table 6: HealthDrive adaptation strategies

5.2.2 SARA (Chronic patients)

The SARA project is intended to provide a user interface for chronic disease patients self-monitoring in the form of a (Windows based) tablet PC. The project is currently evolving to provide multi-device access to the application, via the use of regular Windows desktop computers, Android and iOS tablet devices and smartphones; it is also exploring the possibility of introducing TV-based devices. This project is now in a pre-market phase, after successful field tests using real patients from the Andalusian health system.

Next, as in HealthDrive section, a description of a typical use case is presented, jointly with the adaptation actions which would be carried out supported by Serenoa framework.

*Maria, an elder woman who has a chronic disease, uses daily SARA interface ('Chronic Patients' application) in order to check her medication and doctor's appointments. The avatar is in charge to offer an alternative way of navigation through the system, although Maria is allowed to hide this component at any time, just by clicking or saying it (**Patients' adaptation feedback**). She has some visual problems so the system reads the indications and the graphs or reports which are sent from her doctor (**Visual impairments adaptation**). At the beginning of the interaction the system proposes Maria to measure her blood pressure since it has been requested by her doctor. She starts getting her blood pressure but she doesn't remember well how the procedure is. Several errors happen and then the avatar, with a more empathic attitude (**Error management adaptation**), offers her a demo video which shows how it works. This explanatory material is augmented with links to external content (e.g. Wikipedia) in order to clarify the technical terms or the meaning of certain parts of the process.*

In Table 7 the adaptation strategies from the previous scenarios are described by event, condition and action.

	Event	Condition	Action
Patients' adaptation feedback	The patient hides the avatar (by clicking the corresponding action button)	An avatar is displayed for guiding the interaction	.. Next time she will access, the avatar won't be displayed.
Visual impairments adaptation	Patient starts the interaction with the system	Patient is known to have visual problems	The system turns on the avatar interface and its voice capabilities
Error management adaptation	Patient is asked to follow a procedure	Some errors happen or the access to the procedure is being delayed	A more empathic attitude is configured in the avatar

Table 7: SARA adaptation strategies

5.3 Results

The raw data from our evaluation of the adaptation rules described in section 5.2.1 and 5.2.2 is available in Annex B. In this section we are going to proceed to analyse this information and get some conclusions about the suitability and acceptance of the proposed adaptations.

First, we deal with the HealthDrive scenario where three adaptation rules were presented to participants, namely: language adaptation (i.e. adapt contents to the user's language), avatar adaptation (i.e. degrade the avatar visualization based on device features) and noisy environment adaptation (i.e. switch the main output modality of the system when noise conditions get worse). Please notice that due to the ordinal nature of the data obtained from the Likert items in the questionnaire (see Annex A.1) we are supposed to use descriptive statistics suitable for this kind of data (see (Göb et al., 2007) and (Kislenko & Grevholm, 2008)). For example in Figure 20 we have represented the evolution of the mode (i.e. the number that appears most often in a set of numbers) for the different evaluation criteria, already introduced in section 5.1.2. Next we go to a deeper analysis of each adaptation rule:

- **Language adaptation** rule acceptance is quite constant since almost all values except Confidence are assessed with 4-Agree. Actually participants seem used to this kind of personalization (*"It is a quite standard adaptation and it seems natural to me*").
- **Avatar adaptation** has some noticeable values such as User Performance (2-Disagree) and Future Use (5-Strongly Agree) assessments. It seems that they don't trust on the benefits of this adaptation for carrying out the tasks with fewer difficulties although they would be willing to use the system once they have experienced this adaptation capacity. Some comments point out to a probable loss of capacities due to this adaptation, for example: *"The change is quite disruptive. The loss of capacities from one device to another seems to be very big. It's difficult to say because I haven't tested the final application*", *"I doubt if this transition is not going to bother users. It is strange to pass from a very dynamic character as the desktop version to a quite static one, in the mobile device"*. However there are other users who are comfortable with this new interaction paradigm: "*Whatsapp look-and-feel seems to me very appealing and natural to everybody which has a mobile."* We think it could be the cause that brings to people assesses so good the future use of the mobile application. Besides the question about the consistency across devices also follows a 4-Agree mode.
- **Noisy environment adaptation** mode is around 3-Undecided and 4-Agree except to Timeliness of the adaptation criterion whose mode is 2-Disagree. We collect some comments which could be helpful to clear this point such as: *"Since I have some experience in speech processing I have doubts about how set the threshold to launch this behaviour. For example, what if you are watching TV at home with your mobile? It would be necessary some validation from users"* or *"I doubt about the trigger of this scenario"*.

In summary we could conclude the acceptance and suitability of the adaptation rules, designed for the HealthDrive scenario, seem to be perceived as quite remarkable. It would be necessary to be aware of users' concerns about the loss of capacities when avatar is adapted to mobile devices and the trigger of system's actions based on environmental conditions as subtle as the noise.

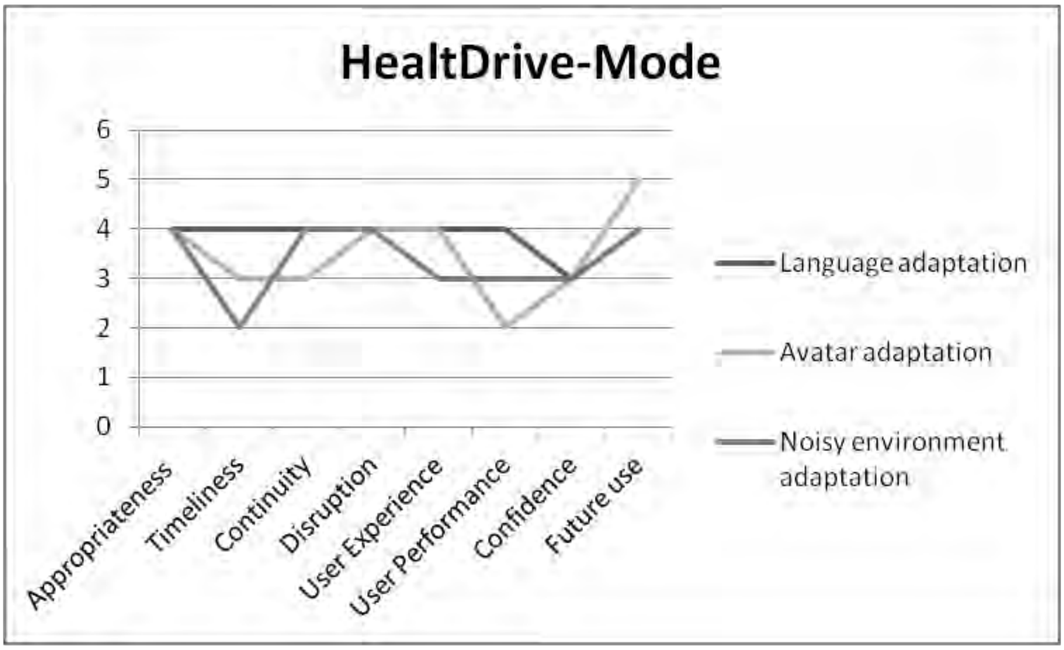


Figure 20: Mode of the different evaluation criteria in HealthDrive scenario

If the punctuations of all the items for each adaptation rule are aggregated and we calculate the means, then we obtain the bar graph depicted in Figure 21. We notice that the language adaptation is the best valued. It might be expected since is the most natural of the set and users recognize to be used to it. The adaptation based on the environmental noise has the lower value although it is clearly over the 3-Undecided anchor.

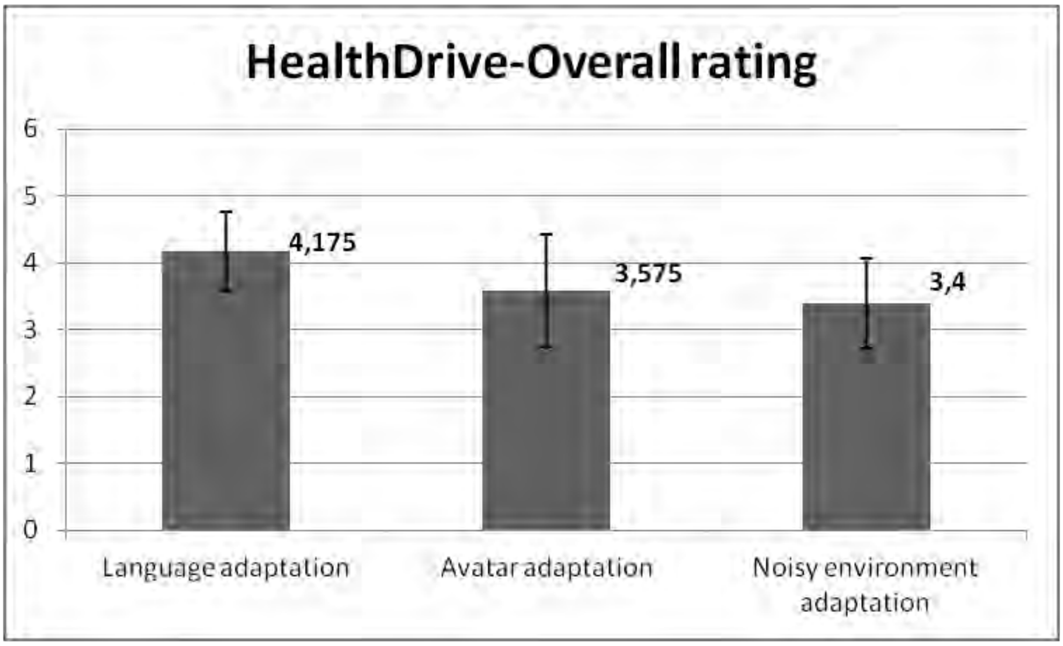


Figure 21: Overall rating for the different adaptation rules in HealthDrive scenario

Regarding SARA use case the adaptation rules are: patient's feedback (i.e. system offers to users the possibility of hiding the avatar and then the user's decision is taken into account to next interactions), visual impairment adaptation (i.e. users who have visual difficulties get the contents via audio) and error management adaptation (i.e. avatar's behaviour is adapted to the user's performance in order to avoid frustration). Figure 22 represents the mode of the evaluation criteria for SARA prototype. Next we go to a deeper analysis of each adaptation rule:

- **Patient's feedback:** mode of all evaluation criteria is 4-Agree. Thus we could conclude that offer to users the possibility to be part of the adaptation process is a good strategy. Participants' comments at this respect are: *"I think it is necessary and it would make that users feel in control"* or *"I like to be free to choose what I want"*.
- **Visual impairment adaptation:** values of the mode are in the same range, between 4-Agree and 5-Strongly Agree. It is remarkable that it is the only adaptation rule which has aroused the deep predisposition (i.e. 5-Strongly Agree) to use in the future the application. Jointly with the high value of appropriateness criterion and some participants' comments such as: *"Accessibility is key for an application"*; we could say this feature is very welcomed.
- **Error management adaptation:** again range of values is between 4-Agree and 5-Strongly Agree. However several issues could be emphasized from the inspection of the raw data (Annex B.2) and the users' comments. Firstly, some distrust to the appropriateness of the avatar's behaviour modification (one of the users assesses with a 2-Disagree the first question). Notice that this part of the evaluation was simulated and the adaptation was only described to users. So maybe we weren't able to convey the aim of the adaptation or it would be necessary a personal experience/interaction. Besides one of the users was a bit scared about the change of personality of the avatar: *"It annoys me since it is not a foreseeable behaviour"*. Secondly, there are some doubts about the suitability of the timeliness to launch this adaptation. It is showed in two punctuations as 3-Undecided in the Timeliness criterion. Again we only specified the adaptation was triggered once some interaction difficulties are detected. For example there was a comment regarding this issue: *"What if I am playing around and system detects I'm failing in my task?"* Surely a more concrete proposal had been more inspirational.

To sum up, SARA prototype with Serenoa adaptation capabilities has been evaluated and users, in general, like the adaptation proposal we have made. On the other hand some concerns about the appropriateness and timeliness of adaptation strategies have been slightly questioned.

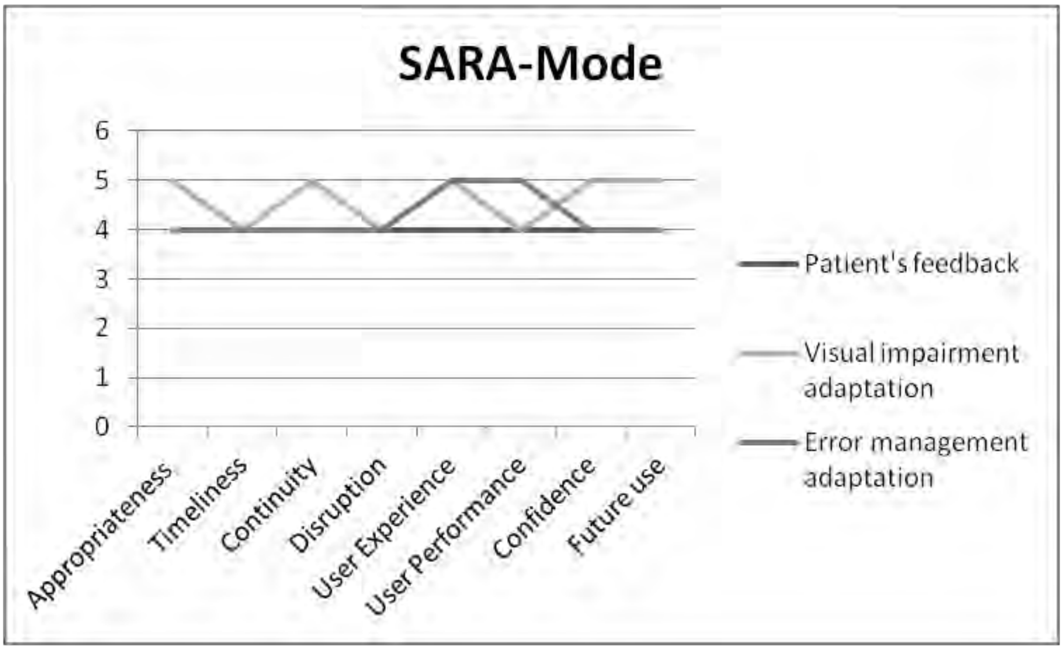


Figure 22: Mode of the different evaluation criteria in SARA scenario

The aggregation of the values of SARA prototype gives us the bar graph displayed in Figure 23. Mean values are quite close and if anything Visual impairment adaptation could be pointed as the better assessed. The focus of this adaptation rule to improve accessibility seems to be very interesting for users.

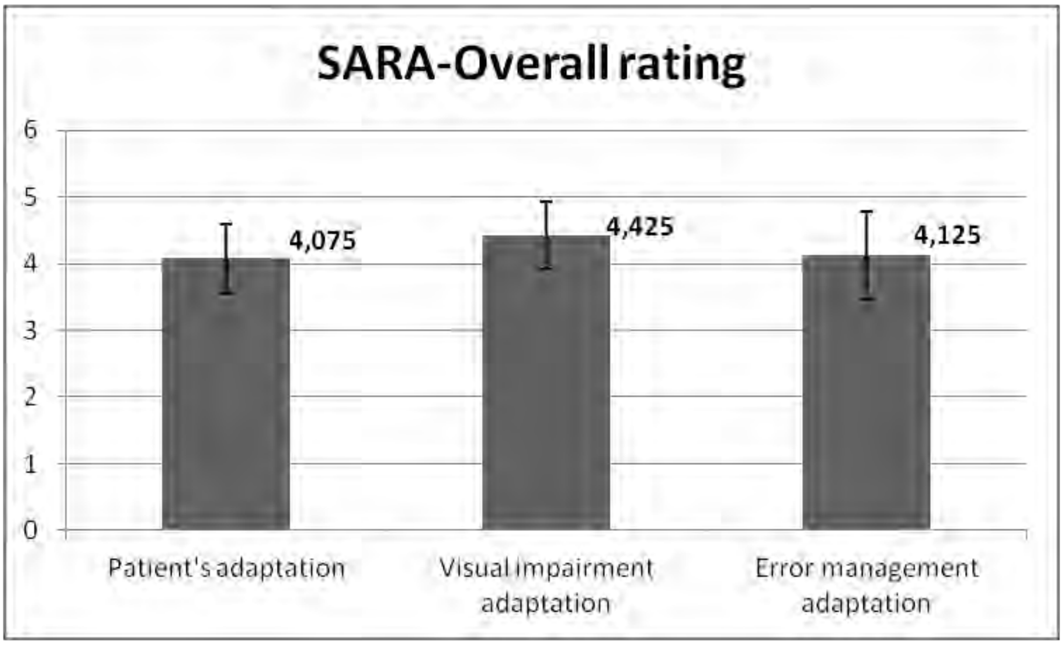


Figure 23: Overall rating for the different adaptation rules in SARA scenario

5.4 Conclusions and Future Work

We could conclude the acceptance and suitability of the adaptation rules designed for the HealthDrive scenario seem to be perceived as quite remarkable. It would be necessary to be aware of users' concerns about the loss of capacities when avatar is adapted to mobile devices and the trigger of system's actions based on environmental conditions as subtle as the noise.

SARA prototype with Serenoa adaptation capabilities has been evaluated and users, in general, like the adaptation proposal we have made. On the other hand some concerns about the appropriateness and timeliness of adaptation strategies have been slightly questioned.

Regarding TID's prototypes (i.e. HealthDrive and SARA) we plan to make a new user testing to evaluate a whole interaction, trying to focus on those features added by the Serenoa framework.

6 Conclusions

6.1 Summary

In this deliverable we have described the results of some initial prototype evaluations that have been conducted within the Serenoa Project.

In particular, the document reports the evaluation carried out at SAP on the adaptive HMD-based prototype, the test conducted by CNR to assess the adaptation of a multimodal UI in the car rental domain, the work done at W4 on evaluation in a business scenario, and the evaluation carried out by TID on two adaptive prototypes (HealthDrive and SARA).

6.2 Future Work

Since we described a first evaluation of the prototypes, the work in this task (T5.3) needs to be continued for the remaining part of the project.

In particular, a second version of this deliverable – namely: D5.3.2 Second Evaluation (users, development tools) is planned for the end of the project. In this second version we plan to further refine and improve the evaluation activities on Serenoa prototypes.

7 References

(Bongartz et al., 2012) Bongartz, S., Jin, Y., Paterno, F., Rett, J., Santoro, C., Spano, L.D. Adaptive User Interfaces for Smart Environments with the Support of Model-based Languages. To be published in Proc. AmI 2012.

(Coutaz et al., 1995) Coutaz, J., Nigay, L., Salber, D., Blandford, A., May, J., Young, R.M.: Four easy pieces for assessing the usability of multimodal interaction: the CARE properties. INTERACT 1995: 115-120

(Göb et al., 2007) Göb, R.; McCollin, C. & Ramalhoto, M. Ordinal methodology in the analysis of Likert scales Quality and Quantity, Springer, 2007, 41, 601-626.

(Kislenko & Grevholm, 2008) Kislenko, K. & Grevholm, B. The likert scale used in research on affect-a short discussion of terminology and appropriate analysing methods 11th International Congress on Mathematical Education, Monterrey, Mexico. Retrieved August, 2008, 17, 2008.

(Paymans et al., 2004) Paymans, T.E., Lindenberg, J., Neerincx, M. Usability trade-offs for adaptive user interfaces: ease of use and learnability. In Proc. IUI '04. ACM Press (2004), 301-303.

(van Velsen et al., 2008) van Velsen, L., van der Geest, T., Klaassen, R., Steehouder, M. User-centered evaluation of adaptive and adaptable systems: A literature review. Knowl. Eng. Rev. 23, 3 (2008), 261-281.

(Serenoa D2.4.1) Paternò, F., Santoro, C., Deliverable 2.4.1 Criteria for the Evaluation of CAA of SFEs, (R1) August 2011

(Serenoa D2.4.2) Paternò, F., Santoro, C., Spano, L.D., Deliverable 2.4.2 Criteria for the Evaluation of CAA of SFEs (R2), August 2012

Acknowledgements

- TELEFÓNICA INVESTIGACIÓN Y DESARROLLO, http://www.tid.es
- UNIVERSITE CATHOLIQUE DE LOUVAIN, http://www.uclouvain.be
- ISTI, http://giove.isti.cnr.it
- SAP AG, http://www.sap.com
- GEIE ERCIM, http://www.ercim.eu
- W4, http://w4global.com
- FUNDACION CTIC http://www.fundacionctic.org

Glossary

- http://www.serenoa-fp7.eu/glossary-of-terms

Annex A. TID's prototype evaluation support material

Annex A.1 Questionnaire

1. The adaptation of the application has been appropriate...

Strongly disagree	Disagree	Undecided	Agree	Strongly agree
□	□	□	□	□

2. The application has changed to user's/context's characteristics at the right moment...

Strongly disagree	Disagree	Undecided	Agree	Strongly agree
□	□	□	□	□

3. After the adaptation happened I went on normally with the task I was carrying out...

Strongly disagree	Disagree	Undecided	Agree	Strongly agree
□	□	□	□	□

4. The adaptation has made the interactive experience more appealing...

Strongly disagree	Disagree	Undecided	Agree	Strongly agree
□	□	□	□	□

5. Thanks to this change I would use the application better...

Strongly disagree	Disagree	Undecided	Agree	Strongly agree
□	□	□	□	□

6. I think this adaptation would help me to use the system without errors...

Strongly disagree	Disagree	Undecided	Agree	Strongly agree
□	□	□	□	□

7. I trust on the system to find out my needs and then, to apply the most suitable adaptation...

Strongly disagree	Disagree	Undecided	Agree	Strongly agree
□	□	□	□	□

8. It was easy to use the mobile version of the service...

Strongly disagree	Disagree	Undecided	Agree	Strongly agree
□	□	□	□	□

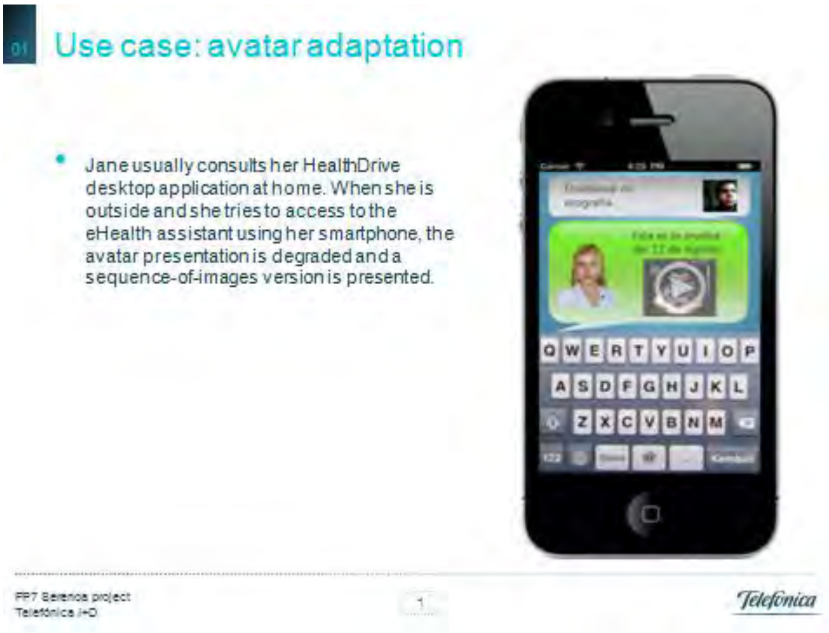
9. I would be likely to use this tool in the future (if needed)...

Strongly disagree	Disagree	Undecided	Agree	Strongly agree
□	□	□	□	□

10. Does the adaptation annoy you?

11. Do you have any additional comments?

Annex A.2 Supporting material



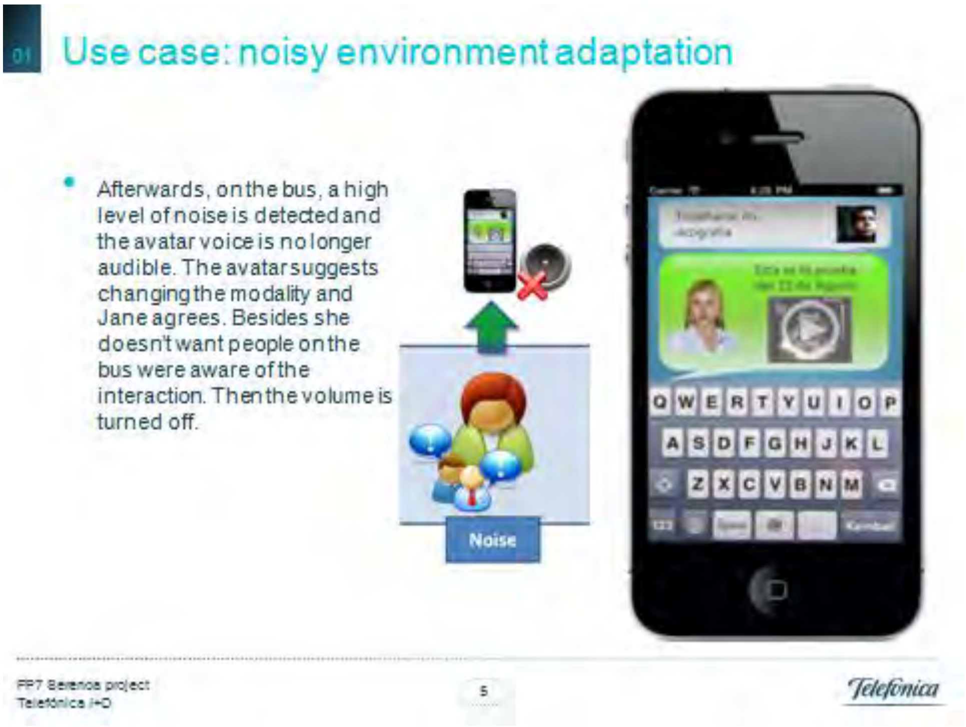
01 Use case: avatar adaptation

- Jane usually consults her HealthDrive desktop application at home. When she is outside and she tries to access to the eHealth assistant using her smartphone, the avatar presentation is degraded and a sequence-of-images version is presented.

FP7 Serenoa project
Telefónica I+D

1

Figure A.1: Motivation slide to avatar adaptation (HealthDrive)



01

Use case: noisy environment adaptation

- Afterwards, on the bus, a high level of noise is detected and the avatar voice is no longer audible. The avatar suggests changing the modality and Jane agrees. Besides she doesn't want people on the bus were aware of the interaction. Then the volume is turned off.

FP7 Serenoa project
Telefónica I+D

5

Telefonica

Figure A.2: Motivation slide to noisy environment adaptation (HealthDrive)

Annex B. TID User testing results

Annex B.1 HealthDrive

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
User1	4	3	4	4	5	5	3	-	4
User2	4	4	4	5	4	4	4	-	4
User3	4	5	5	4	4	4	5	-	5
User4	5	5	4	5	4	4	4	-	5
User5	4	4	3	4	4	4	3	-	4

Table B.1: Language adaptation scores

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
User1	3	3	2	3	4	2	3	4	4
User2	3	3	3	4	4	4	3	4	5
User3	4	4	3	4	4	5	4	4	5
User4	4	3	3	4	5	4	3	3	5
User5	4	4	2	3	4	2	3	3	4

Table B.2: Avatar adaptation

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
User1	4	2	4	3	2	3	3	-	3
User2	5	3	4	3	3	4	3	-	4
User3	4	2	4	4	3	3	3	-	3
User4	4	4	4	4	3	3	3	-	4
User5	4	3	4	4	3	3	3	-	4

Table B.3: Noisy environment adaptation

Annex B.2 SARA (Chronic patients)

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
User1	4	3	4	4	4	4	4	-	4
User2	4	4	4	4	5	4	5	-	4
User3	4	4	4	3	4	3	4	-	4
User4	5	5	4	4	4	4	4	-	5
User5	4	5	4	3	4	4	5	-	4

Table B.4: Patient’s adaptation scores

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
User1	4	4	4	4	4	4	4	-	4
User2	5	4	4	5	5	4	4	-	5
User3	5	4	5	4	5	4	5	-	4
User4	4	4	5	4	5	4	5	-	5
User5	5	4	5	4	5	4	5	-	5

Table B.5: Visual impairment adaptation scores

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
User1	5	4	4	4	5	5	4	-	5
User2	4	4	4	4	5	5	3	-	4
User3	2	3	4	4	4	4	4	-	4
User4	4	3	4	4	5	5	5	-	5
User5	4	4	4	4	4	4	4	-	4

Table B.6: Error management adaptation scores