Deliverable 2.3.2

Project Title	**Next-Generation Hybrid Broadcast Broadband**
Project Acronym	**HBB-NEXT**
Call Identifier	**FP7-ICT-2011-7**
Starting Date	**01.10.2011**
End Date	**31.03.2014**
Contract no.	**287848**
Deliverable no.	**2.3.2**
Deliverable Name	**2nd Report on User Validation Results**
Work package	**2**
Nature	**Report**
Dissemination	**Public**
Author	**Jeroen Vanattenhoven (KU Leuven), Björn Stockleben (RBB)**
Contributors	**Jennifer Müller, Bettina Heidkamp-Tchegloff (RBB), David Geerts (KU Leuven)**
Due Date	**31.03.2013**
Actual Delivery Date	**18.04.2013**

Table of Contents

1. Introduction

The work of Task 2.3 evaluates prototypes and working demos with users. These user evaluations happen in three phases throughout the project. This report describes the results of User Validation Phase 2. Whereas the previous phase of evaluations was conducted with paper prototypes and one clickable prototype, this phase makes use of the first working demonstrators namely: The Smart App Store and the Settings Application. In addition, further planned extensions to the Settings Application (profiles and second device connection) were tested using paper prototypes. Available results were communicated to all partners at the Consortium Meeting at THM in March 2013.

The second user validation phase is intended to evaluate usability and user experience aspects of the chosen designs using advanced mock-ups and working prototypes in a mix of qualitative and quantitative methods. Additionally, conceptual paper prototype testing is applied to provide further functional enhancements to the original designs.

Three laboratory tests were performed. The first test is an extensive evaluation of the Smart App Store. The second test analyses the performance of the refined settings application and the third test assesses possible approaches to multi-user, device connection and profile handling extensions to the settings application.

The test results comprise specific recommendations for the improvement of the tested applications as well as generalized insights for SmartTV application development.

By 31.6.2013, this deliverable will be enhanced by an annex to cover user experience testing of a first early field trial performed by RBB in May 2013 and a working prototype of the group recommendation engine by KU Leuven.

2. Test Nr 1: Smart App Store (KU Leuven)

Figure 1. The home screen of the Smart App Store

2.1. Approach

This section describes the user evaluation of the Smart App Store by NEC (see Figure 1). We conducted a lab test with the demonstrator with 11 participants between 19 and 47 years old. More background information about the participants was gathered via a questionnaire during recruitment. We will describe this information in the following paragraphs.

There was a mix of singles, people living with their partner and people living with their family (see Figure 2). Seven participants were female; four were male.

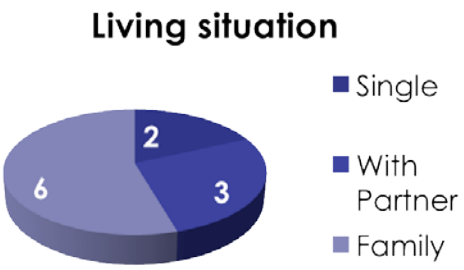


Figure 2. Participants' living situation

We also inquired into the TV products that participants owned, and what functionality they used. What we can see on Figure 3 is that most participants are using quite a lot of interactive TV functionality, except for recommendations. Two out of eleven participants had satellite TV.

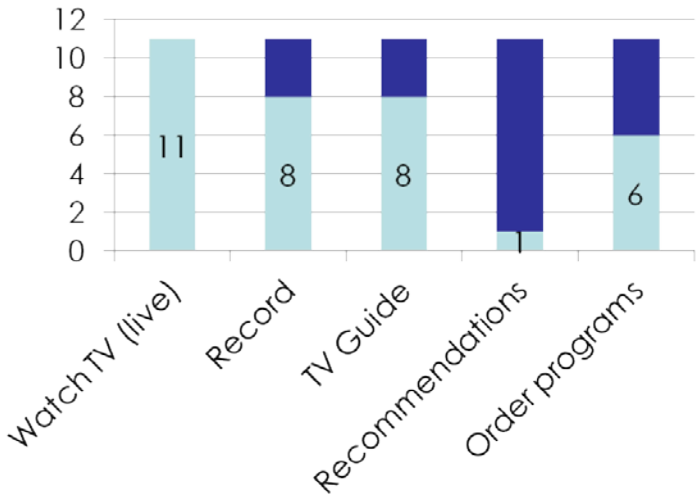


Figure 3. Interactive TV functionalities used by participants.

It was also useful to know if certain participants already owned a Smart TV and what functionality they used (see Figure 4). The majority did not own a Smart TV. The most used Smart TV functionality of the participants that did own such a TV was watching Internet videos.

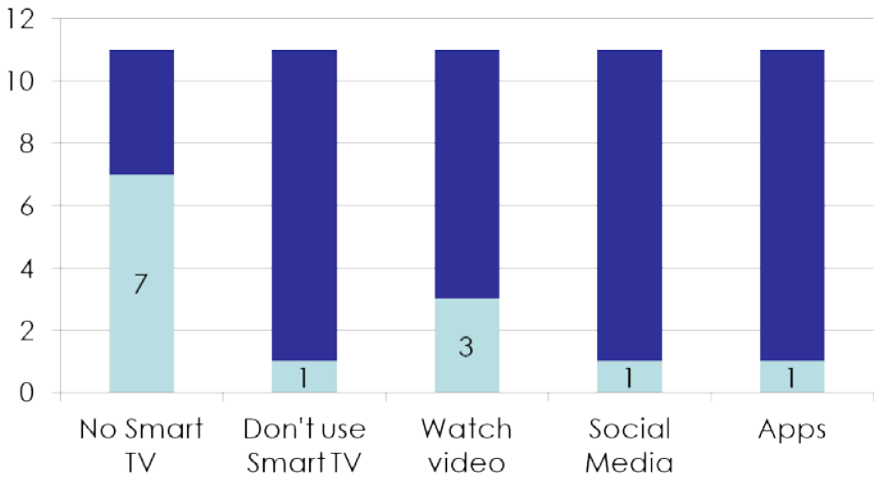


Figure 4. Smart TV ownership and used functionalities.

Finally, we also wanted to know which app stores they currently used (see Figure 5), and on which devices they used this (see Figure 6). Most popular was the Apple app store; the most used device was the smart phone.

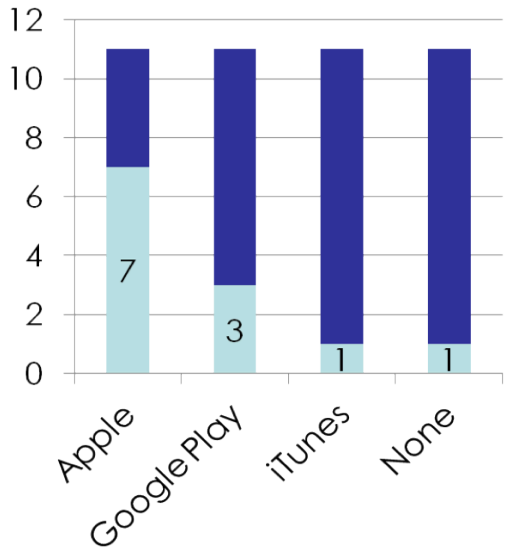


Figure 5. App stores used by our participants.

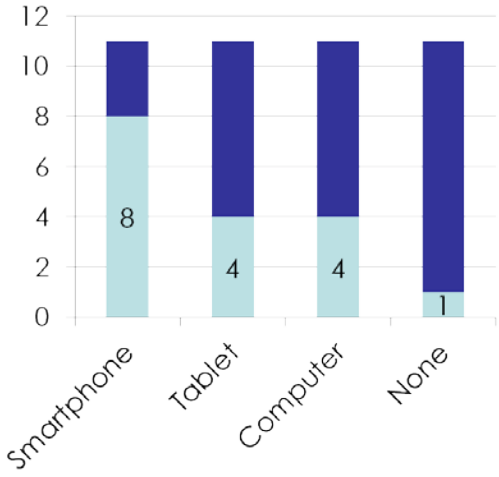


Figure 6. Devices on which participants use an app store.

For the setup we opened the demonstrator on a browser on a MacBook, which was then connected to the TV screen in the lab (See Figure 8 and Figure 9). After evaluating a number of tasks with the demonstrator, participants also explored the Xbox App Store in order to get a clearer view on the differences between the two regarding content, interaction and user experience.

For the demonstrator mouse and keyboard were provided (since it was a web-based demo), for the Xbox the XBox remote control was provided (See Figure 7). For better readability of the text on TV, we used the browser to increase the App Store's font size.

Figure 7. The remote control that belongs to the Xbox.

Figure 8. Setup in the lab

Figure 9. Setup in the lab.

2.2. Test Flow

Participants were first welcomed and introduced to the goal of the evaluation. Then we asked them to perform three basic tasks with the demonstrator. Each task was observed and logged using the usability logging software. The following **tasks** were included:

1. Create a new account and log in.
2. Look at the information describing an app, and download this app. Find this app after it has been downloaded.
3. Provide your feedback and comments on an app.

After each task participants were provided with a set of questions. The first three questions related to the IBM ASQ **subjective usability scale** [1]:

1. Overall, I am satisfied with the ease of completing this task.
2. Overall, I am satisfied with the amount of time it took to complete this task.
3. Overall, I am satisfied with the support information (on-line help, messages, documentation) when completing this task.

All three statements need to be scored on a 7-point Likert scale from "strongly agree" (1) to "strongly disagree" (7). Each statement provides insight into a different aspect of usability. The IBM ASQ subjective usability score is then calculated as the average of the scores on the three statements [1]. After giving the scores, participants were also asked to describe in their own words how this task went: what went good, did they encounter any problems?

Then, participants were asked to **explore the Xbox App Store** (browsing items, installing, provide ratings…) for approximately five minutes in order to make clear the difference between the two, and making it easier for them to provide feedback. After this exploration phase researcher and participant discussed their experiences based on the following **questions**:

1. How does the Xbox app store compare to the prototype? On which points is it better/worse for you?
2. Which app categories do you need? Which are not necessary?
3. How do you find interesting apps (via friends, family, newspaper reviews…)? Which information does an app need in its description (file size, ratings…)?
4. Which kinds of apps would you like to have on your Smart TV?
5. Did you understand "reputation"?
6. Do you sometimes provide feedback on apps? When, when not? Why, why not?

2.3. Results of test Nr 1: Smart App Store

In this section we will present and discuss the results of the user evaluation.

2.3.1. Task 1: Create an account and log in

For the first task (create an account and log in), participants were a bit confused with the word "authenticate" on the home page (see Figure 10): Nine out of eleven participants had trouble knowing where to start in order to create their account. P1 (participant 1) said: "*At first I didn't know where to go to log in. 'Authenticate' was not clear enough for creating an account. Once I discovered that it went smooth*".

SUGGESTED SOLUTION: replace the Authenticate button by a Register button, and a Login button.

Five participants did not really understand why they had to indicate their preferences at the account creation page (see Figure 11), as illustrated by P2: "*The purpose of the ranking system is not clear: cost or app. A lot of stars = expensive of cheap?*" P7 stated: "*I did have some problems filling in the preferences. This was not clearly described and given the fact that I had never used this app store I couldn't have filed this in correctly.*"

SUGGESTED SOLUTION: separate the preferences section from the registration form, and clearly indicate what each preference means, which end of the star rating is positive, and which is negative. This will have to be evaluated again with users.

Finally, after logging in they did not always realize that they had logged in. This was due to an intermediate window that appeared after logging in. The message "*you are logged in as …*" was only shown when this intermediate window was closed (see Figure 12). This problem was mainly observed as all participants were waiting and looking around at the intermediate window, before trying to close that window. After this they noticed that they were logged in. Only one participant, P4, illustrated this in the comments on this task: "*It would also be nice if you would be logged in immediately after you have created your account*".

Figure 10. The term "authenticate" was not clear.

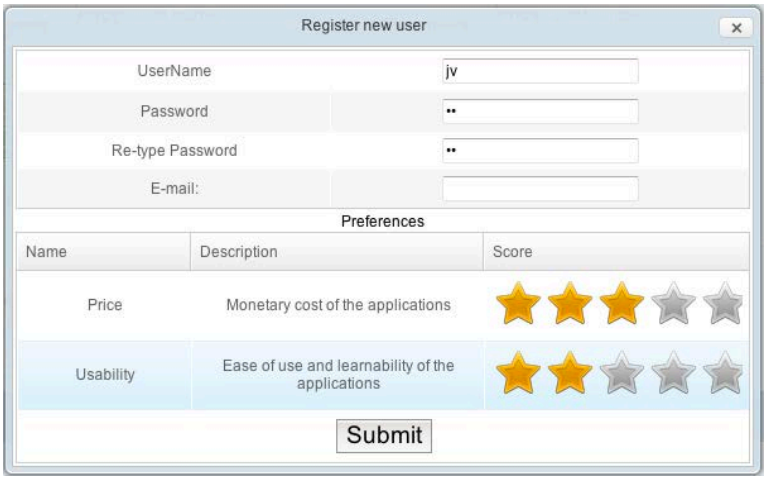


Figure 11. Indicating preferences at account creation.

Figure 12. Intermediate window "User Authentication" is shown after the user logs in. Notice the message "you are not logged in" in the top right corner.

Figure 13. After closing the „User Authentication“ window, it is clear the user „jva“ is logged in.

On the positive side: Overall, there were no severe usability issues in this task. Every participant was able to complete this task. Participants appreciated the short registration form required to create an account (see Figure 11), as illustrated by P3: "*Filling in the data went smooth. Only requires the most essential data*". This refers to other real-life registration forms, which often require too much information; it can take quite some time to complete such forms.

The **subjective usability score** is **3**, which is the average of:

- Ease of use: 3
- Time needed: 2
- Support information: 4

When looking at the subjective usability scores, ease of use and time needed seem quite realistic: every participant was able to complete the tasks, but there were some difficulties at the start of the registration process, and knowing when they were actually logged in. Three participants described this task as "*smooth*". For support information, things are more nuanced: P5 commented: "I did not need supporting information for this, so I can't judge this". But given the problems with the preferences – participants did not really know why they were there – more or better support information might be necessary.

So, on the one hand participants lacked the need of support information, on the other hand they might have needed it in some cases, but they might not have realized this need when answering the questions. This can explain the quite average score of 4 on this task.

2.3.2. Task 2: Look at an app's information, download it, and find this app.

A first issue for task 2 was that participants did not notice that the downloading had started, nor that it had finished (see Figure 14 before installing; see Figure 15 after installation). After waiting for some time all participants just assumed the app was downloaded and moved on with the task. No progress bars were shown, not status messages indicating that the download had started or finished.

Five participants referred to this issue in their comments, as P4 illustrates quite clearly: "*There was insufficient information when the app was installed. There was no feedback, which made me doubt whether I had actually installed the app at all. I also had to search a little bit where to click in order to download the app. I was clicking on the logo at first.*"

SUGGESTED SOLUTION: Provide a notification for the start, progress (progress bar), and end of the download process. At the end also inform the participant where the app has been downloaded.

Secondly, one participant indicated that besides the metadata there was no information describing the app. P11 states: "*Not really clear what the app entails (what you can do, how do I use it…)*".

SUGGESTED SOLUTION: add a brief description to the app information that explains what it is, and what you can do with it.

Finally, because participants were using mouse and keyboard for the demonstrator they complained that they could not click the logo or title of the app in order to open it. They could only use the "view more" link (see Figure 16). Two participants reported this problem as illustrated by P1: "*I wanted to open the app by clicking on the name. That didn't work. The same happened with the image*".

SUGGESTED SOLUTION: This issue probably has its origin in the fact that the prototype was still a website, and not yet converted to a TV app. When converting it into an app for TV, care has to be taken that the design fits the use of a remote control. On the homepage view of an item, the whole item should be clickable, and should respond to input from the remote control.

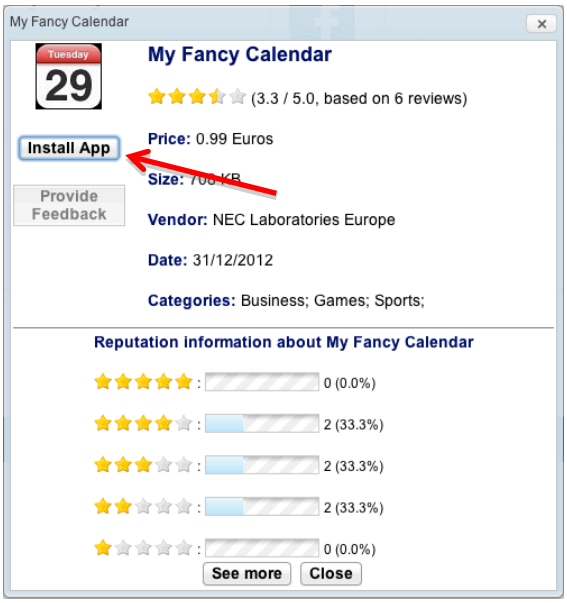


Figure 14. The basic app information appears when the user opens an app.

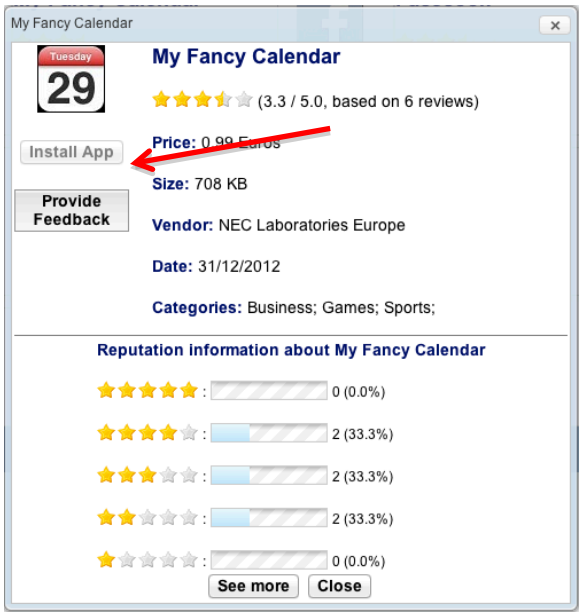


Figure 15. The app has been downloaded and installed, but the only thing different with respect to Figure 14 is that now one button is greyed out, and the other is active.

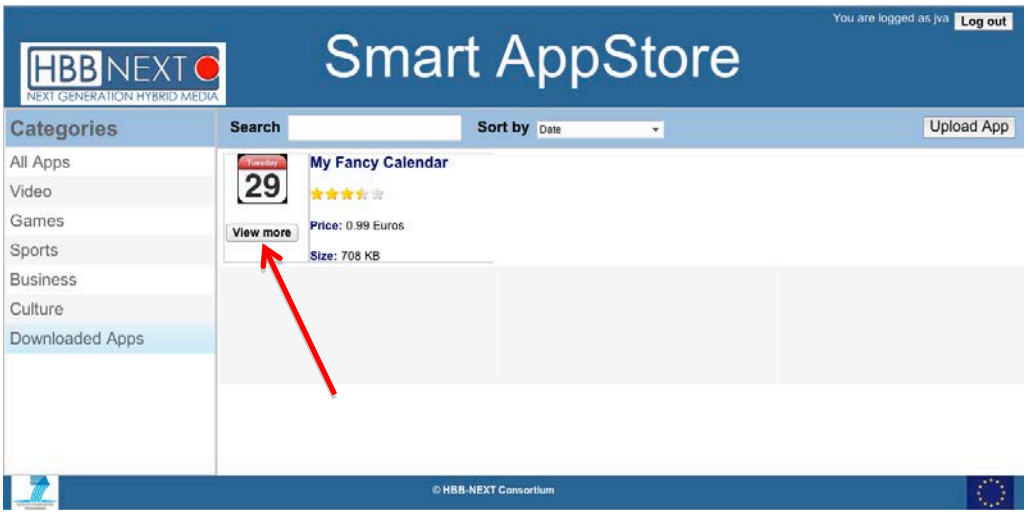


Figure 16. The downloaded app is found in "Downloaded Apps". Only „View more" could be clicked to open an app.

On the positive side, all participants were able to complete this task. Furthermore, all participants found the installed app in "Downloaded Apps" quite fast (see Figure 16). Only one participant experience problems with this; P9 says: "*Finding the app (after installation) was more difficult for me because I didn't know if I had to stay in the app store or close it*". This issue has already been reported and refers to the fact that there is no feedback on the download process.

The **subjective usability score** is **2.33**, which is the average of:

- Ease of use: 2
- Time needed: 2
- Support information: 3

These scores look quite good for ease of use and time needed. No severe issues were found. Participants mainly had difficulties where to click the app in order to open it, and one participant experienced a problem finding the app after installation. Six participants described this task as "smooth"; P2: "*It did go smooth though in general*"; P3: "*Very easy to find more information and download. 'Downloaded items' is a clear term.*" The score for support information is a little bit better than in Task 1. One participant, the same participant as in Task 1, explained that she did not need support information; P5: "*again, I did not need supporting information*". P4 provided the worst rating of all participants, a seven, and commented: "*There was insufficient information when the app was installed. There was no feedback which made me doubt whether I actually installed the app at all.*"

This indicates that this user and possibly other users did misunderstand the question. Support information entails tooltip support, help and documentation. It is different from the information available in the interface of the task itself such as the menus, labels, and buttons. In case the user interface for a certain task is not clear, and does not point out how to achieve something, then support information should be available for the user such as FAQ, help etc. In her comment the user clearly refers to the lack of feedback on the download process, which is information that should be present in the interface itself, and cannot be considered support information. To conclude: Though the support information received good scores, ten participants do not mention it in their comments, which might indicate that they didn't need it (as P5 indicated). The fact remains that the prototype contains no support information at all and that the scores for the support information should be taken with a grain of salt.

SUGGESTED SOLUTION: Although adding more information on a TV user interface is quite hard because every user interface element on TV has to be substantially larger than on the web, we would recommend to include some basic support information on the TV app.

2.3.3. Task 3: Open an app, provide feedback or comments on this app..

We first have to note that five out of eleven participants were not able to execute this task because of technical difficulties. Therefore, these results are based on the data from six participants.

The biggest and only issue in this task was that no participant found how they should provide textual feedback on an app. See Figure 17 and Figure 18). P7 illustrates: "*It was very easy, but how I could provide comments such as the ones I saw, wasn't clear*". P9 continues: "*The task went smooth. I wanted to write something besides giving a rating, but I couldn't find how to do this*".

SUGGESTED SOLUTION: immediately provide the comment box below the ratings. As this window contains very little information or elements, this is certainly possible.

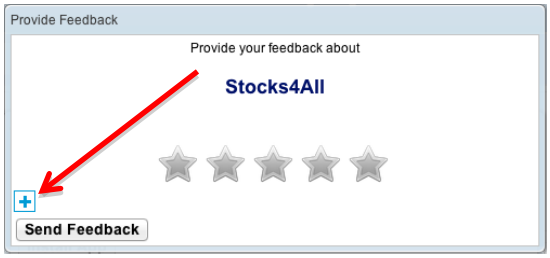


Figure 17. The + button had to be clicked in order to open the text field for written feedback.

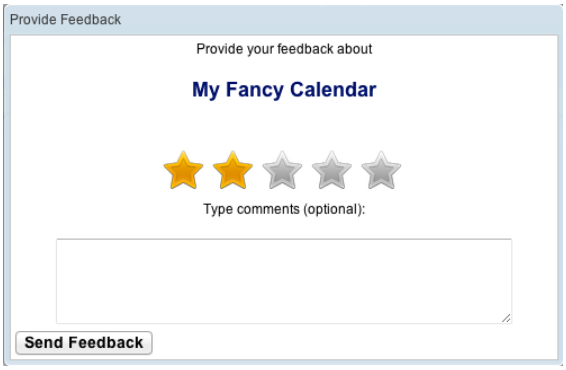


Figure 18. This is the window with the text field; no participant found this.

For the remainder of this (small) task, no participant experienced problems.

The **subjective usability score** is **1.33**, which is the average of:

- Ease of use: 1
- Time needed: 1
- Support information: 2

The subjective usability scores in Task 3 are even better. But we have to take a closer look to interpret these results. The main reason for this is that no participant was able to provide written comments. At the same time these participants gave excellent scores to this task. Therefore, we have to nuance the subjective usability score, as half the functionality was not found. The excellent score can perhaps be explained by the fact that now participants needed very little time to complete this task, since they only found the ratings.

2.3.4. How does the Xbox App Store compare to the Smart App Store?

In order to retrieve more feedback from participants, it is always useful to provide an alternative for comparison. That way people find it easier to formulate feedback than when they only judge or use one application. Therefore, participants were asked to explore the Xbox app store for approximately five minutes, and try out downloading, rating, and navigating the Xbox app store in general. There were no specific tasks for the Xbox, no concrete measurement. This exploration only served as contrast with the prototype in order to make it easier for participants to provide feedback. A measurement on itself would have little use since the Xbox is a finished product, and the Smart App Store is a prototype. Furthermore, the Xbox is ready for TV while the Smart App Store is a website, currently designed for computer screens. Finally, the Xbox contains real applications already attributed to menu categories, while the Smart App Store contains dummy apps, which are not organized with regard to specific menu categories. We are not interested therefore in the measurement on itself, but in how the Xbox does things differently, and why that is better or worse for the users.

What participants found positive for the Xbox app store is that the interaction felt more intuitive and responsive. The Xbox was described by participants as intuitive (2x), useful (1x), clear (1x), interactive (1x) and responsive (1x). This is partly due to the fact that the Xbox is a finished product, and that we evaluated a prototype. In addition, for the Xbox they could use a remote control, and the Xbox has an interface specifically designed for TV and for remote control. Nonetheless, this is still very interesting to know. Participants for example appreciated the animations that occur when going from one section (ex. “video”) to the next section (ex. “music”). P1 illustrates this point: “*The Xbox is more intuitive. With a few button presses you can move along*”. P4 continues: “*The system is more responsive and when you change screen, it happens in a smooth motion, and that’s more pleasant to watch*”.

CONCLUSION: It is important to provide a usable application. But it is also important to make the interaction responsive and intuitive. This is more about how the user feels about the interaction. In addition, the Xbox user interface was also found visually appealing. Six participants mentioned this and used the words: inviting, pleasant to watch, pretty and eye candy.

P6 says: "*The Xbox app store is really attractive*". P10 comments: "*There is more eye candy in the Xbox store*".

CONCLUSION: This point illustrates that providing a visually attractive design has a positive effect on the user experience.

Two participants indicated that the Xbox provided better feedback on the downloading process: the start, the progress and the end of the download were clearly communicated to the user. P9 describes: "*The Xbox is better for downloading: you first get a screen to ask if you really want to install it, then you get a screen where you can see the download*".

CONCLUSION: This point has been reported before in depth: provide feedback to the users about what the system is doing. In this case, we are talking about the downloading process.

While the previous paragraphs point out the positive findings of the Xbox app store, there are also a number of issues on which the Smart App Store does better. With regard to the Smart App Store, participants found the navigation easier. Five participants mentioned this in their feedback. If one starts at the home screen of the Xbox and navigates deeper into its structure, only the current section is visible. When participants then navigated a few levels down inside the hierarchy, they found it difficult to orient themselves inside the Xbox. On the prototype, the navigation was always present at the left side of the screen. P1 clarifies: "*I do miss some structure. Where am I inside the Xbox? How do I return Home? On the prototype you could see the menu items on the left, so you knew where you were*". P9 adds: "*The prototype is better for switching menus*".

CONCLUSION: When converting the Smart App Store from a computer-oriented design to a TV-oriented design, make sure that there are not too many hierarchies in the TV app. Furthermore, we recommend showing where the user is located inside the hierarchy. If the hierarchy is not too extensive, one way of showing this is the breadcrumb (for example: HBBTV > App Store > Facecook).

Furthermore, the Smart app store contained user reviews while the Xbox did not. Three participants noted this: "*There is no extra information on the apps; you don't know what they do right away, and you don't see any comments*" (P7).

P10 continued: "*The prototype store provides better information on reviews and ratings. The prototype store also provides more detailed information about the app such as the date of the app*". Participants also appreciated the extra information in the form of metadata such as file size, vendor etc.

CONCLUSION: Participants orient themselves on these reviews to judge an app. Therefore, the user reviews form essential information supporting purchases.

On the Smart app store, all participants knew where to find the downloaded apps. On the Xbox this was not the case, despite the proper feedback on the download process itself. Downloaded apps on the Xbox are located in a separate section called "My Apps". Two participants commented on this. P3 said: "*The segment 'Downloaded App' for the prototype was useful*".

CONCLUSION: The 'Downloaded Apps' section was very clear to the users.

Finally, participants also found it easier to search on the Smart App Store, but this was most likely due to the available input technique. On the Xbox they had to use the remote control to compose keywords in the search box, which was very cumbersome. This was also reported previously in D6.2 where we compared the different interaction techniques – remote control, speech, and gestures - on Microsoft's Kinect. P5 illustrates this: "*What I don't like on the Xbox: typing. Something takes forever because you have to traverse the whole alphabet, if you want to go back, you need to use 'back' and you can just go left the whole time. The back button is reasonably small.*"

CONCLUSION: Due to the fact that this issue has its origin in the interaction technique, we refer to the insight we gained in D6.2. Voice interaction is more suited for search tasks on TV. For character input with the remote control, the Xbox showed an alphabet on one line. This means that for going from A to Z 25 button presses are needed. Other, more efficient ways for providing character input should be explored.

2.3.5. Which App categories are (not) useful?

The Xbox was found to have a very clear organization of categories. Only one issue was found here: the difference between the "TV" and "Video" categories was not clear to participants. Three participants commented on this, for example P1: "*The difference between TV and Video needs to be clear though*". It is not entirely clear what the category TV apps would contain, since you have TV channels on every TV. The Xbox contained on TV app, from Sky Television. The app allows you to watch the Sky TV channel and the Sky TV guide. The question is whether an app is the right place to provide this functionality.

CONCLUSION: The main categories are: Video, Music, Games, and Social. More research is probably needed for other categories.

Participants would also like a category "most popular", which present a list of all apps, but sorted on popularity. With regard to the categories "business", "culture", "learning/education", "sports" there was no clear indication as to participants' preferences. Consider "sports" for example: you could have sport games, and sport videos. In which category should these then be posted: sports, games, or videos?

CONCLUSION: At this point it is not clear whether you should an app only in one category, or place a sports game both in sports and games. What we can conclude is that when a choice is made, that this choice has to be applied consistently. Furthermore, a list of all apps is also a required feature. This has to be sorted; one way of sorting these apps is via popularity. Other ways can be added (for example: newest) but care has to be taken not to provide too many different sorting mechanisms.

One observation made during evaluation is that participants seem to understand the categories themselves not only based on the category name, but also in the items located inside.

CONCLUSION: Careful placement of apps inside the right categories is equally important as finding the right term.

2.3.6. How do you find interesting apps?

We asked participants how they currently found interesting apps. This mostly meant apps on Smartphone as can be seen in Figure 6. Eight out of eleven participants received interesting suggestions from their (real-life) social network such as friends, families and colleagues. Almost the same amount of people relied on review in their newspapers and magazines. Less used are recommendations and reviews in the app store itself, as can be seen in Figure 19.

CONCLUSION: We should try to incorporate recommendations from friends via the App Store, and allow users to consult their newspaper and magazine reviews. This is not that straightforward since many stakeholders are involved.

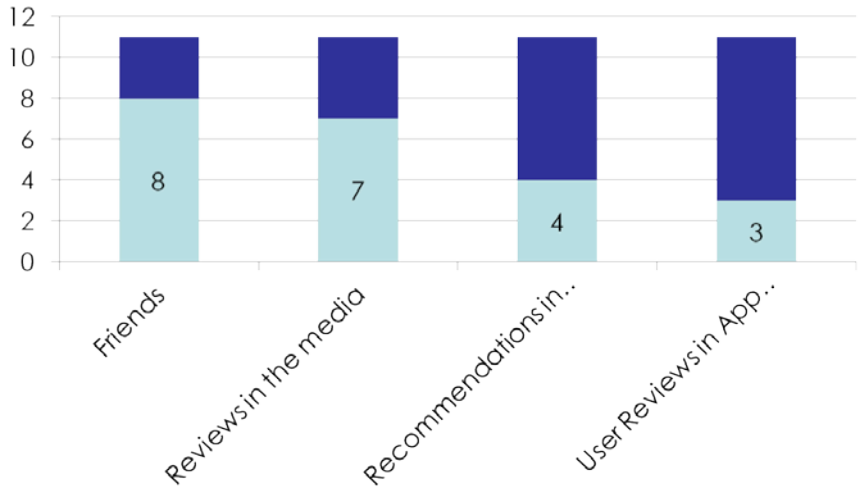


Figure 19. Sources for interesting apps.

2.3.7. Which information do you need to judge an app?

In the app store itself users can normally find information describing the app. The Xbox mostly contained a brief but clear description on the app. This was absent in the Smart App Store. On the other hand, the Smart App Store contained some metadata related to the app (see Figure 20). Of this information, the price, the user ratings and the user comments (when asked), were found useful. File size, date and vendor were found less useful overall. Though, we have to note that some participants, with a more technical background, would like to see the file size, and vendor information. The first is required to judge if there is enough room available for it to be installed on the Smart TV.

The latter is required to judge the reputation of the vendor. This can occur when the participant has already installed other apps by the same vendor.

Figure 20. Basic app information, showing metadata such as file size and price.

2.3.8. Which apps would you like to see on your Smart TV?

We were interested in finding useful suggestions for apps that could be developed on Smart TV. This resulted in a range of possibilities: shopping apps, YouTube, Facebook, Skype, apps with background info on TV programs, newsfeeds, the weather etc. However, we did not find a clear trend or direction in these answers. This is probably due to the fact that Smart TVs are still a very new concept, which makes it difficult for people to imagine what they would find valuable. This stresses the importance of testing and trying out of working demonstrators as close as possible to the real-life situation of the consumer. This is what we will carry out in User Validation Phase 3 of the project.

2.3.9. Did you understand “reputation”?

A significant part of the window describing the app on the Smart App Store was attributed to a “reputation” section as shown in Figure 21. This showed the distribution of user ratings. All participants understood how this was constructed or calculated. However, not all participants found that it was necessary information. The overall score (based on the aggregated user ratings) is more important in this respect.

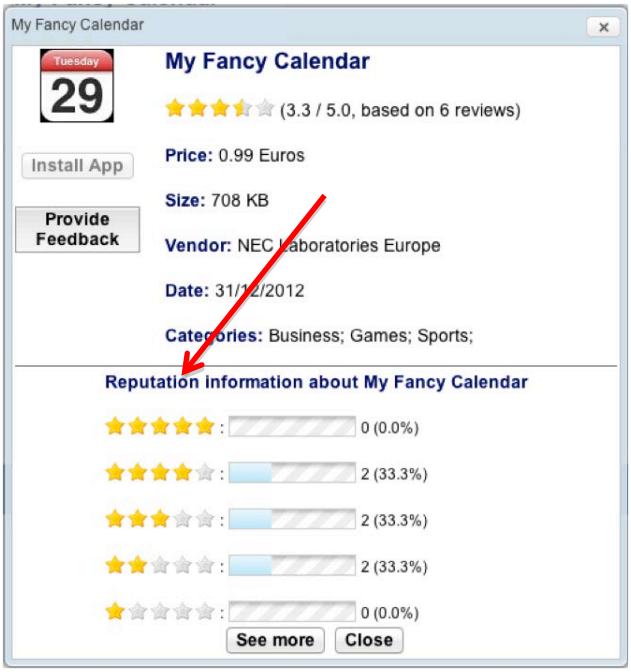


Figure 21. Reputation section about an app.

2.3.10. When do you give ratings? Why, why not?

Participants do not always give ratings. Seven participants never do this. The reason why participants don't provide ratings is that they don't use other peoples rating either; therefore they probably do not see the relevance of providing feedback. These people rely entirely on recommendations from friends or newspaper reviews, and do not consider the user ratings. For the people that provide feedback, four insist on providing written feedback in order to explain their rating. These participants mostly provide ratings for very good or very bad apps. The motivation for providing ratings is to help other people judge apps, and/or to punish or reward the maker of the app.

2.4. List of Recommended Solutions

Use terminology that is clear to the user

The word “authenticate” for example, was not clear to the user as the place for creating accounts and logging in. “Reputation” was another example of terminology that was not clear.

Provide a clear and simple structure for the app categories

Organizing all available apps into a clear navigation structure is not an easy task. For now, we have seen that the following categories are suitable for an app store: video, music, games, and social. Furthermore, a list of all apps would be useful, but preferably in an ordered way such as “most popular” apps. A number of other categories (business, learning etc.) could be added but this depends more and more on personal taste. And one has to make sure that there are not too many categories (see next point). Card Sorting sessions with end-users can provide further clarity as this method is especially suited for finding out how to structure information from the user’s perspective [3]. Note that the categories themselves are important, and that the apps inside each category are equally important for the user.

Design the UI for TV, not for the Web (computer)

On TV this exercise is even more difficult since every UI element such as text and buttons have to be a lot larger than on a computer. After all, the user is further away from the screen and still has to be able to use the device. The implications are that less UI elements and text can be shown on TV. With regard to the App Store the text should be significantly larger, menu items should be larger, and some elements on the detailed app information can probably not be displayed in one window. Finally, the demonstrator will also have to provide remote control interaction as it evolves from a working prototype on the web to a more realistic TV app.

Communicate the system status to the user

An important guideline in Usability Design is communicating the feedback of the system to the user [2]. This was not the case for the download process, and the login status. For the download process it is important to indicate the start, the progress and the end of the download. With regard to the login status it is important to show this immediately when a user has logged in.

Support different ways to find interesting apps

Our results, though with a limited amount of people, show that users have different ways of finding their way to interesting apps. People rely on their social connections that recommend interesting apps. Furthermore, the newspapers and magazine they are subscribe to, also review different apps. Social connections and media review therefore formed two important sources of finding interesting apps. Future research could first confirm these results, and second, investigate ways to incorporate these sources into the Smart TV interface.

Provide the right information on each app

We evaluated the information that describes an app, because at this stage many forms of information can be offered, but it is not clear which one is useful. Furthermore, as the design for TV uses larger dimensions than for computer, less can be shown. We found that the essential information is the average user rating, a basic description, the price, and the user comments. The latter however takes quite a lot of space; therefore, we recommend showing this optionally (on a 2nd level). File size could also be shown optionally. Care has to be taken how to show this. The average user does not work with Mbytes or Kbytes. Neither does he/she know how much space is available. Therefore, the size of the app should be shown with respect to the entire space and available space on the Smart TV.

Use a brief registration procedure

The registration procedure on the prototype was very brief, and participants appreciated this. Normally, registration procedures require a lot of information, often too much from the perspective of the user. Therefore, we recommend to only asking for the essential information. During registration, preference information was asked relating to price and usability. Participants did not understand this; this will require a better explanation.

In addition, this information should probably be located elsewhere; not in the registration procedure.

Provide a pleasant user experience

Participants described the interaction with the Xbox as pleasant, fluent and very responsive. The Xbox design is visually attractive, and also responds very fast to button presses on the remote control. Furthermore, animations are shown when moving from one category to the next. All these elements contributed, perhaps not to the usability, but certainly to the very positive user experience (relating to the interaction itself).

3. Test Nr 2: Improved Subtitles & Sign-language (RBB)

Subtitles and sign language are features of the HBB-NEXT settings application, which has been improved and extended according to the results of the mock-up evaluation documented in D2.3.1.

3.1. Pre-Test

The tests were conducted with a group of 5 hard-of-hearing and 5 deaf persons. The group was aged 35-65. For the test evaluation, we decided not to separate the results by age or gender groups. According to the principles of universal design [4] products should be designed in a way that includes the whole potential user group. Including age or gender increases the risk of a biased evaluation of issues. The test group answered a pre-test questionnaire to assess media consumption habits, technical skills and traits of opinion leadership. The basic insights were:

- All users own a TV and watch TV daily.
- One user does not own a computer, smartphone or a tablet. Three own a computer, one a computer and a smartphone. Five own all three devices. The Internet is used on all of these devices with the exception of two smartphone.
- Only two people own a Smart TV that is actually connected to the Internet.
- Seven users use the Internet on a daily basis, while two state they use it several times a week.
- The Internet is mainly used for communication and information. Six people used video and/or music frequently.
- When it comes to web 2.0 it is striking that four out of five deaf test persons use it in some way, while none of the hard-of-hearing does. This might be partly explained by a minor age difference, but it also supports the general expression from this and other tests sessions that deaf people do embrace Internet technology more intensely.

Technical skills and opinion leadership

We used two scales each to identify *technical lead users* (Hippel, 1986, [9]) and *accessibility multiplicators*. Multiplicators in the sense of this study are such persons, which do have a high reputation in the deaf and hard-of-hearing communities due to their position, knowledge or personality. For their identification we relied on the research on opinion leaders in sociology. The rationale is that the opinion of multiplicator personalities might play a crucial role in promoting the use of the application in the respective hard-of-hearing and deaf communities. The identification of multiplicators follows the opinion leader concept in sociology (Schenk 2007, pp. 350, [8]).

Technical lead users were identified as follows:

Scale 1: How often do others ask you about technical matters (TV and internet)? – (frequently, often, rarely, never)

Scale 2: How often do you ask others for help about technical matters (TV and internet)? – (frequently, often, rarely, never)

We defined those users as lead users who at least answered *often* in scale 1 and stated the same or a lesser frequence in scale 2. In other words, lead users are users which give advice often or frequently and ask for advice themselves in equal or less amount.

Example: scale 1: *often* and scale 2: *rarely* would be defined as lead user, while scale 1: *often* and scale 2: *frequently* would not be counted.

With this definition, we find **3 out of 5 deaf** users being **technical lead users** and **2 out of 5 hard-of-hearing** users.

Accessibility multiplicators were identified as follows:

Scale 1: How often do others ask you about accessibility and inclusion matters? – (frequently, often, rarely, never)

Scale 2: How often you ask others about accessibility and inclusion matters? – (frequently, often, rarely, never)

The discrimination was done in the same way as with the technical lead users, identifying **4 out of 5 hard-of-hearing** users and **2 out of 5 deaf** users as **multiplicators**.

Furthermore, it is important to note that all users had at least a very basic of experience with the general interface logic from prior user tests.

3.2. Approach

The test was conducted with a working application prototype. The user was seated in front of the TV set and interacted with the TV remote control. We used task-based testing. In order to generate additional quantitative data, for each task we recorded the execution time and the error level [5]. The error levels indicate how severe each usability issue is:

- **No problem** – The user fulfilled the task without any problems at all.
- **Minor problems** – The user got briefly stuck, but was able to understand and correct her errors without help.
- **Major problems** – The user experienced problems that could only be solved with external help and that would have likely led to failure in a normal usage situation.
- **Failure** – The user could not solve the task, even with help from the moderator.

Participants were encouraged to think aloud. We were aware that this has an impact on the execution time, but felt that the application was not yet mature enough to neglect this source of information fully. The analysis was based on a categorized evaluation of the test protocols, identifying and grouping related issues of multiple users.

3.3. Test Flow

3.3.1. Tasks

1. Activate the settings application, activate and adjust the subtitles to your liking and return to the TV programme.
2. Now go back to the settings application and change the subtitle language.
3. Activate the settings application, activate and adjust the sign language interpreter to your liking and return to the TV programme. (Task was limited to deaf testers only)

During the task all actions and comments were observed and logged. After the task the tester was asked if she/he had any additional comments.

3.3.2. Prototype application screenshot

The **main menu** of the settings application looks as follows (on-screen texts for tests were written in German).

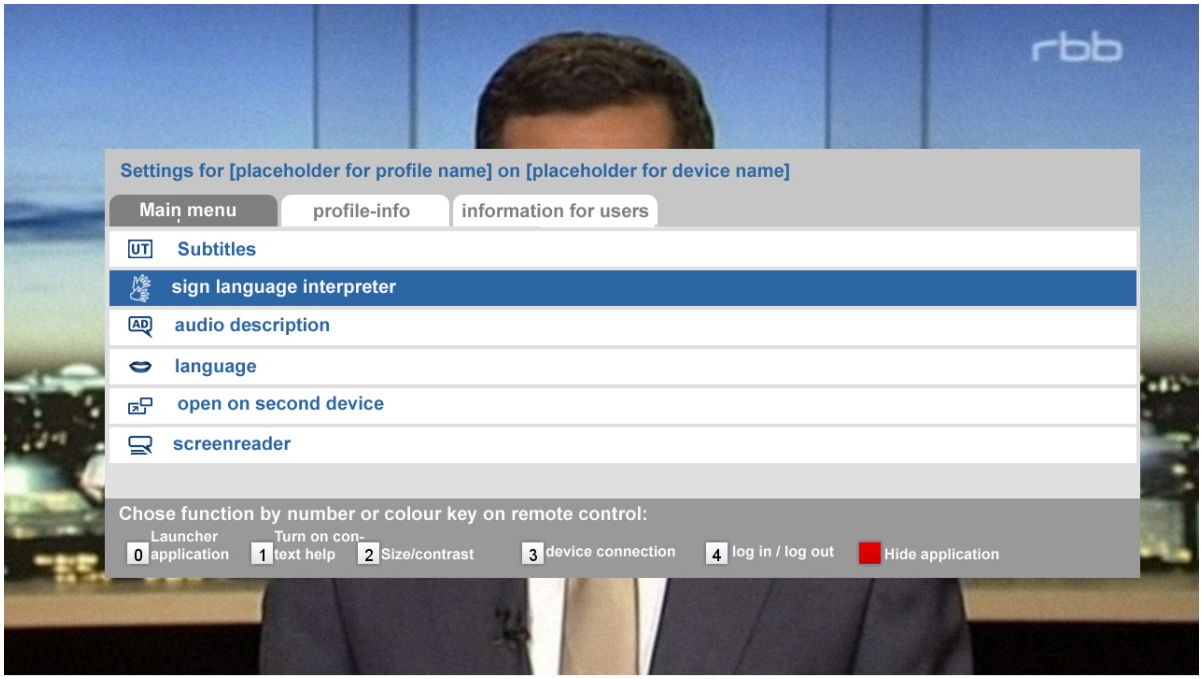


Figure 22. Main menu of settings application

3.4. Results test Nr 2: Improved Subtitles & Sign-language

3.4.1. Execution times and error rates

The error rate for **task 1** was *no problem* for 6 users (2 hard-of-hearing, 4 deaf users) and *minor problems* for 4 users (3 hard-of-hearing, 1 deaf user). Thus all users were able to figure out the use of the application themselves, which is a satisfying result. However, a look at the execution time reveals a large variation within the test group:

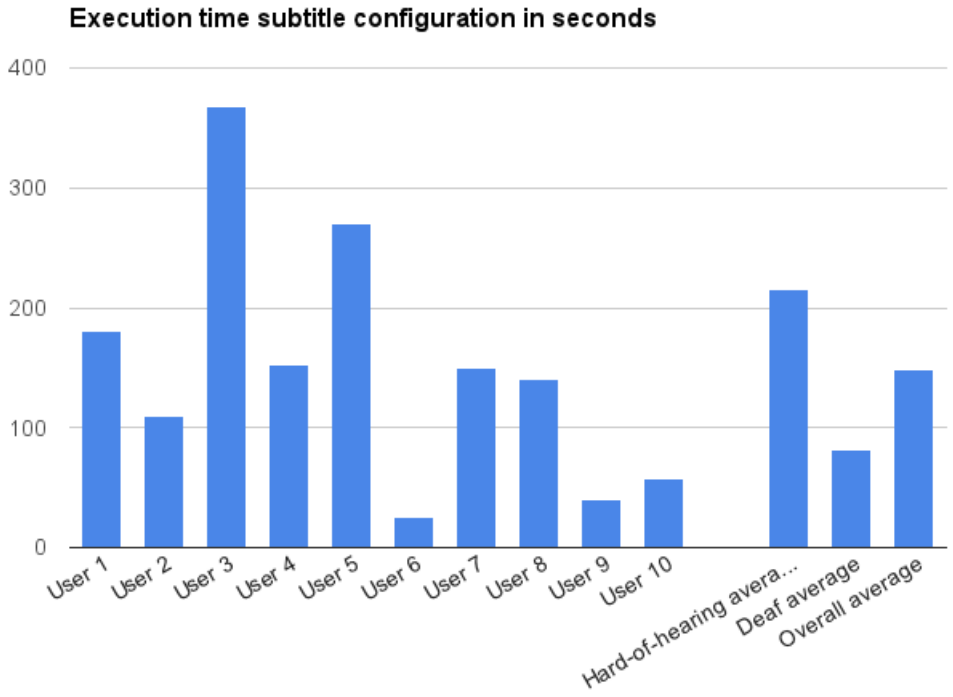


Figure 23 Task 1 - Execution times

The average execution time among the deaf users settles at 82 seconds compared to an average of 216 seconds for the hard-of-hearing. Though a certain variation in the execution time influenced by the degree of familiarity with digital interfaces, it can be doubted that some of the users would have completed the task in the living-room situation. Missing several minutes of the desired programme is not acceptable. Especially first-time users could be deterred from the start, driving them back to using traditional teletext subtitles where still available. These results indicate a clear need to straighten out further usability issues as identified in the qualitative part of the study.

In **Task 2** the participants were asked to change the language of the subtitles. This task appeared to be fairly easy, as all users had already spotted the subtitle language feature in the task before. All users remembered the position of the subtitle language option and consequently the error rate was at *no problems* throughout the test group. With an average execution time of about 22 seconds and no mentionable deviations, the comparison of the execution times indicates that the subtitle language option has been correctly placed into the navigation structure.

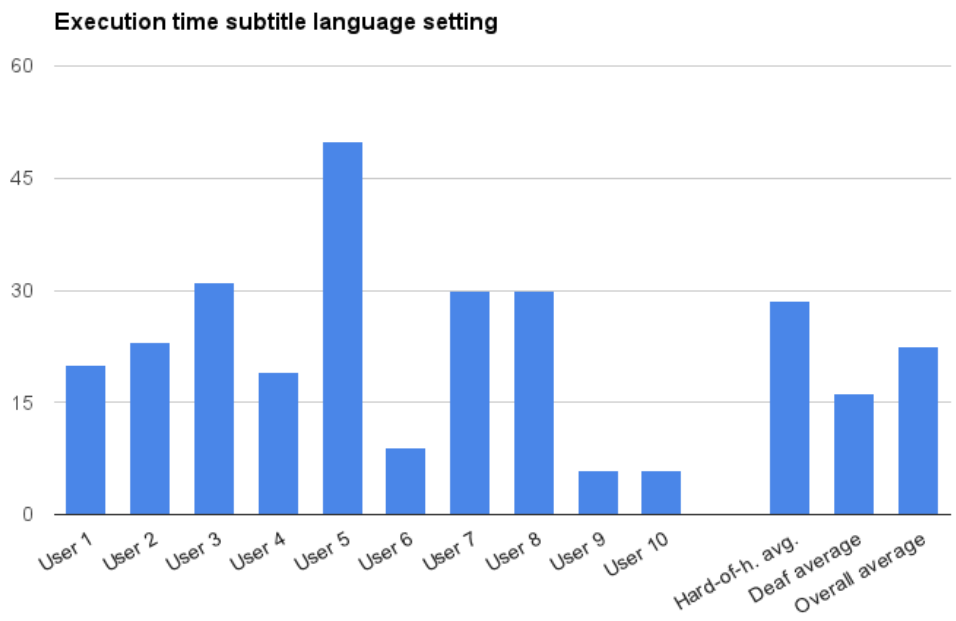


Figure 24 Task 2 – Execution times

Task 3 asked from the deaf participants to activate and configure the sign language interpreter. User 5 was added to the test, as she speaks sign language and asked to do the task as well. Three users had an error rate of no problems, the other three had minor problems and thus all users were able to solve the task without external help. A look at the execution time table shows an average time of 86 seconds. To soften the effect of the one exception at user 7 with 185 second, we calculated as well the median, which was at 72 seconds. In general these execution times are reasonable, given that the users performed a full configuration of all available settings for the sign language interpreter. However the exception of user 7 should not be neglected and the causes should be investigated in the qualitative results.

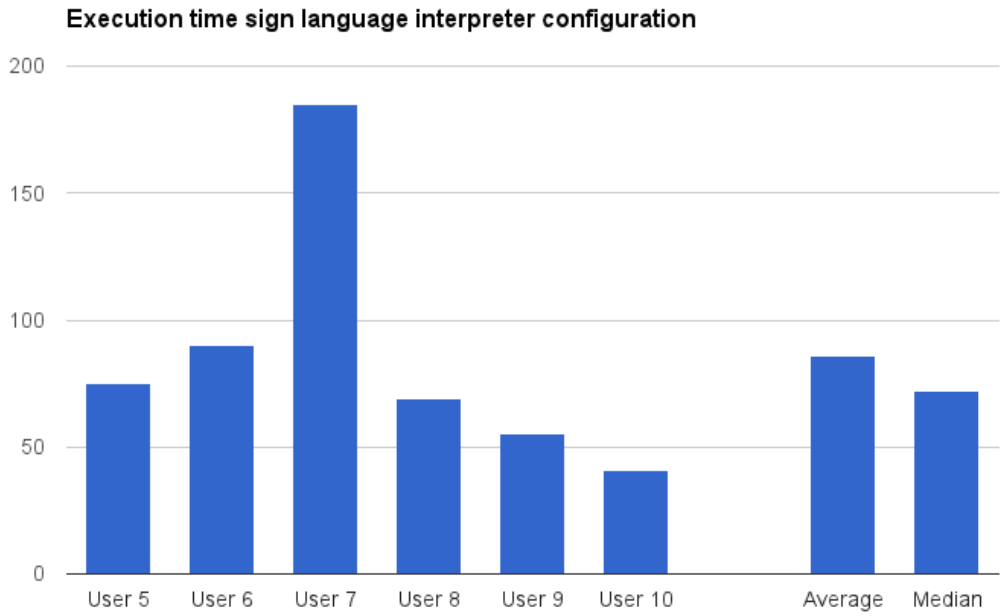


Figure 25 Task 3 - Execution times

3.4.2. Protocol evaluation

A protocol analysis entails that we review all user trajectories, incorporate their actions and comments, in order to identify the causes of the usability issues. We identified a number of issues, which are listed here in condensed form translated from the original German protocol.

Note: Missing numbers indicate that issues were merged into existing ones due to similarity.

No.	Description & possible solution	Users	Category
1	User selects “settings” in the application launcher bar, but hesitates to push the button. He is unsure whether accessibility services are really subsumed under “settings”. **Solution**: Change “settings” to a more specific name like “settings & accessibility” or introduce a separate entry for “accessibility”.	2	Information architecture

No.	Description & possible solution	Users	Category
2	When starting the settings application, the users push the OK-button too long, so that the application immediately jumps to the first submenu on start. **Solution**: Give feedback that a keystroke was received, even though application loading needs time. Resolve the related technical issues, e.g. clearing key input cache on application start.	3	Feedback
3	User wants to go back to TV picture and usually knows he had to push the red button, but is irritated by unknown feature "device connection". **Solution**: Unknown features can irritate otherwise well-known usage flows. Change place in information architecture and use clear wording.	1	Unknown features
4	User looks at entries in launcher bar, stating: "Mediatheque does mean music, doesn't it?" **Solution**: TV applications have a very wide audience and seemingly established terms might be unfamiliar to part of the audience. Use concrete terms instead of fancy, but artificial words.	1	Wording
5	The launcher bar is shown and the user wants to press red button again to get into the settings application. The user does not fully understand the principle of the red button. **Solution**: Introduce consistent use of the red button.	1	Mapping
7	User presses "0" to hide the settings application, but activates the launcher bar this way. The correct button "red" button is not consistently associated with show/hide of applications. **Solution**: Use key mapping consistently throughout applications.	4	Mapping

No.	Description & possible solution	Users	Category
8	User reads “auf zwei Geräten öffnen” (open on two devices) instead of “auf zweitem Gerät öffnen” (open on second device). User seems to read rather what they expect than what they see, especially with long and unknown words. This holds true even more for elderly users who also have an impaired vision to some degree. **Solution**: Increase readability and use clear and familiar language.	1	Accessibility
9	In subtitle-menu user reads “Umrahmung” (frame) instead of “Umrandung” (border / black edge). **Solution**: Increase readability, especially contrast.	1	Accessibility
10	User choses “frame around TV picture” as subtitle rendering option and dislikes the aesthetics. **Solution**: Not necessary, but it might turn out that this design is not needed.	1	Needs
11	User thinks “language” tab in subtitle menu is about audio. **Solution**: Make more explicit that it is about subtitle language by adding information.	1	Information Architecture
13	The contrast in the bottom menu of the settings application is badly readable. **Solution**: Use stronger contrasts and strokes.	2	Accessibility
14	User tries to move focus onto tab “background”, but has already moved to the tab using the “left” and “right” buttons. **Solution**: The selected tab should be clearly marked, right now it is painted in dark grey. Stick to common design patterns (i.e. tab navigation).	3	Navigation

No.	Description & possible solution	Users	Category
16	User opens subtitle menu, activates subtitles and leaves the settings application without making further adjustments. She did not realize there were tabs with more features. **Solution**: Apply design pattern "tab navigation" in a proper way. (see Figure 27. Incorrect implementation of tab navigation pattern and Figure 28. Correct implementation of tab navigation)	1	Design Patterns
18	User does not know when the configuration dialog has ended. He demands some button or feedback like "press okay to finish". **Solution**: Especially at the end of a task flow users have a need to explicitly finish the task. We might see a clash between TV and PC logic here: On the PC users are used to a lot of confirmation and "save now" dialogues, TV usually accepts any settings instantly.	1	Feedback
19	When user leaves setting application by pressing "0" and hides the appearing launcher bar with "red", pressing "red" again will show the settings applications instead of the launcher bar. **Solution**: Implement consistent behaviour of red button and "0" throughout all applications.	1	Consistency
21	User searches for sign language options under "language" tab of subtitle sub-menu. This is related to issue 2. **Solution**: Introduce clear information architecture, including position information like breadcrumb display.	1	Information architecture

No.	Description & possible solution	Users	Category
22	User does not notice the button “back to main menu”, as it looks like a normal list element. **Solution**: Use visual cues to separate and emphasize button.	3	Grouping
24	Tester is asked to activate the sign language interpreter and says: “Do I have to turn off the subtitles first?” **Solution**: Maybe communicate clearer what is possible and what not.	1	Mental Model

3.5. List of Recommended Solutions

Accessibility

Use clear contrasts and fonts that have at least two dots wide lines. Dark gray - light gray contrasts look nice in visual design, but are not accessible when it comes to reading interfaces. In the tests, visually impaired testers criticized especially the menu at the bottom of the settings application screen.

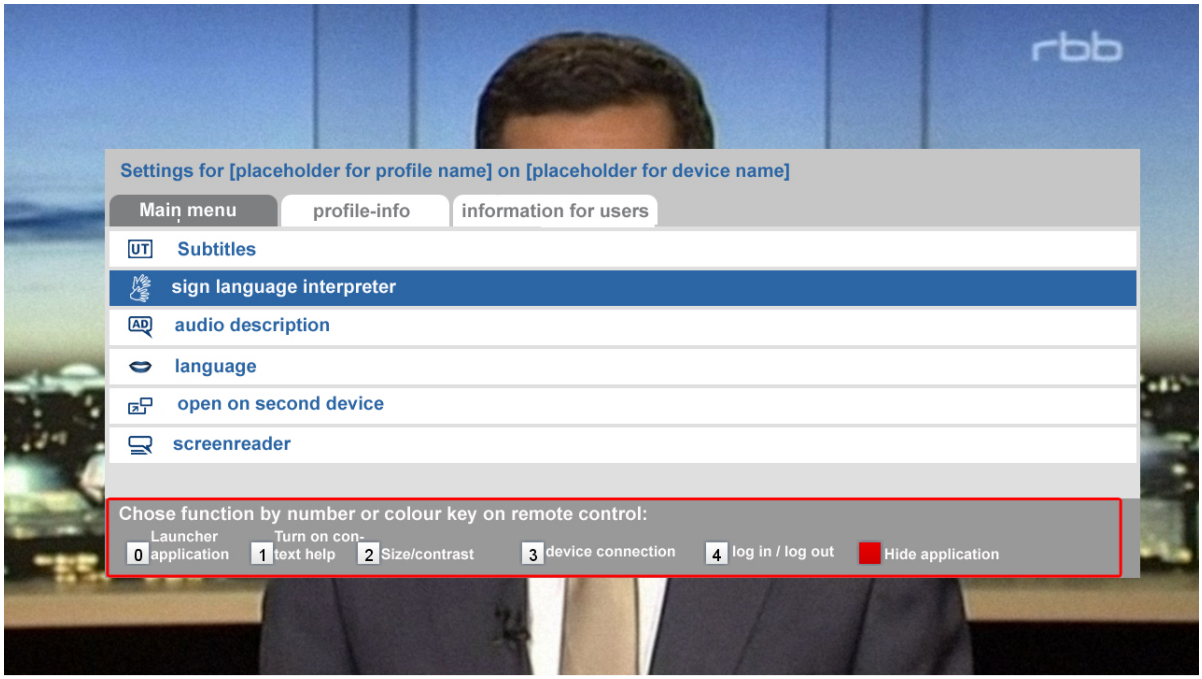


Figure 26. Low contrast in the bottom menu (marked by red box)

Consistency

Seek for consistent mapping of actions to colour buttons and numbers. A re-mapping will most likely fail; any features specific for a particular screen only should not be mapped to keys, but be accessed via menus.

Correct application of established Design Patterns

Established Design Patterns such as tab navigation should be checked for suitability to TV and remote control interaction and either be fully implemented or substituted by a more suitable pattern. An incorrect or inconsistent implementation of patterns leads to mismatches with user expectations and thus frustration.

The two screenshots exemplify the difference: In the correct version, the user clearly sees which tab heading describes the currently active tab, while the inactive tabs are contrasted in grey.

Figure 27. Incorrect implementation of tab navigation pattern

Figure 28. Correct implementation of tab navigation

Feedback

Any more or less reasonable user interaction should yield a feedback. If users cannot identify a clear focus on the screen they are likely to try one of the cursor keys. Thus the focus should always move when a cursor key is pressed and if the layout allows there should be a rolling focus in menus, i.e. the focus should jump to the first item of the menu again when pressing down or right (depending on menu orientation) on the last item. Furthermore, especially TV applications may react delayed in some cases. It has to be made sure that additional key presses are not queued up and cause uncontrolled actions when an application or screen finally shows. The latter is more a technical implementation issue that causes usability problems. All configurations should have a visual preview if applicable. E.g. subtitle configuration should provide a preview even if subtitles are currently turned off.

Logical Grouping

The Law of Proximity (Goldstein 2010, p.106, [6]) should be considered when designing interfaces, i.e. different functional blocks should have a visible additional distance from each other. The Law of Proximity and other Gestalt principles describe the way humans perceive objects and are helpful to determine which UI elements the user will visually perceive as an entity. E.g. a menu entry called “back to main menu” should always have an additional blank space in contrast to the other lines of the menu entries.

Other visual cues to separate functional blocks according to the Gestalt principles are delimiters and a change of background colour.

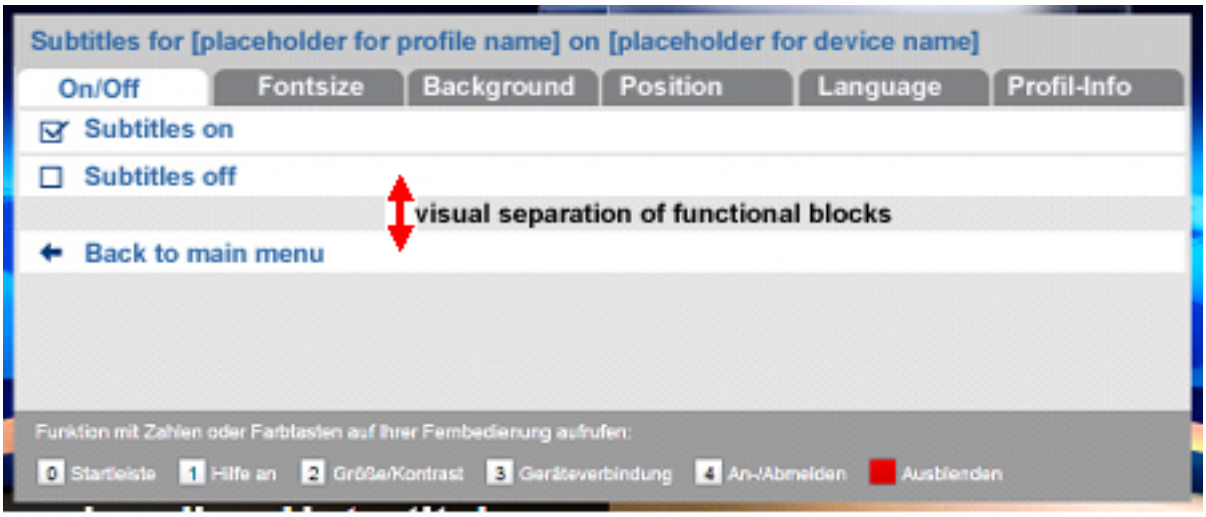


Figure 29. Example of visual separation of functional blocks

Information Architecture

When creating hierarchical information structures, categories should be chosen deliberately to avoid names that are too generic to the user. We found out that “Settings” is not considered a natural top category for subtitling services. Subtitles and other accessibility features are considered services in their own right. Thus it either could be called “Settings & Accessibility” or “Accessibility” could be a category next to “Settings”.

The users do not seem to distinguish clearly between the start bar and the settings application. They tend to see “Settings” as a feature of the start bar. It should be considered that the user’s view on the application often differs from the underlying implementation logic. The design model should always be oriented towards the user, not the implementation model.

Instructions

Users do not read on-screen instructions fully. Instead they loosely assess whether the text somehow matches existing concepts they already know, without checking details. In other words, users read what they expect rather than what is written. This as well holds true when legibility is bad. In the worst case the instruction-problem may turn into a dilemma:

Experienced users do not read instructions, novice users do not understand the instructions. This dilemma can only be resolved if different solutions are sought for the two user groups. There is no “one-size-fits-all” in user instructions.

However when a single solution has to be applied for whatever reason, it should focus on the novice users in the case of SmartTV applications.

Irritations by unfamiliar features

If an application is extended by unfamiliar features, their mere existence can influence tasks that would otherwise be totally clear to the user. Unfamiliar features are likely to draw the attention of the user.

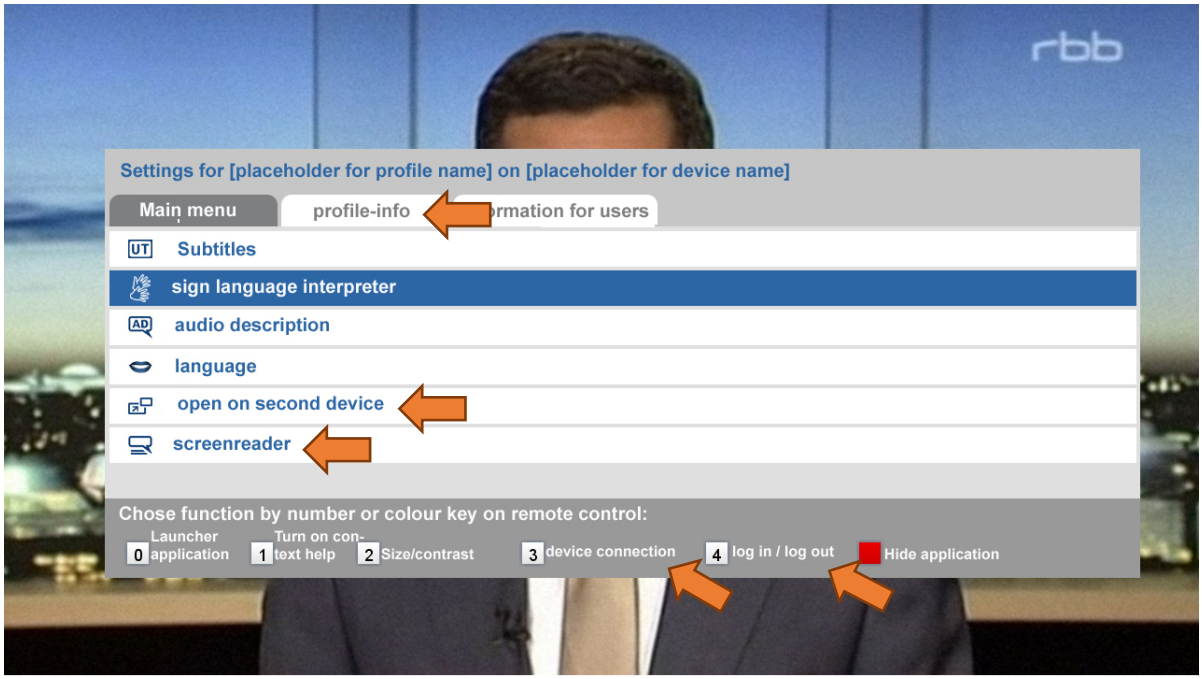


Figure 30. Interface features unfamiliar or new to most users

Expectations due to existing Mental Models

Users may have prior experiences that make them assume things to be possible or impossible. E.g. some users assumed that it would not be possible to show both sign language interpreter and subtitles simultaneously. In addition, a lot of users assumed that only one person at a time can be logged in on the TV-Set.

Wording

Sometimes users have the right flow of actions in mind, but use different terms, so that they do not see matching interaction possibilities in the interface. Even terms that are seemingly familiar nowadays such as “link” or “URL” may pose problems in certain target groups and should be replaced by more “speaking” terms, e.g. “internet address”.

Card Sorting sessions might help to identify the correct wording, and to establish the correct information architecture [3].

4. Test Nr 3: Device Coupling & Multiple Profiles (RBB)

4.1. Approach

The second part of the settings application test was designed to research the possible integration of new features into the application. While device coupling and handling of multiple profiles are relatively straightforward in technical implementation, they constitute new design concept unfamiliar especially to an elderly audience. To assess needs, possible interaction concepts and different mental models evoked by both device coupling and profile handling, we decided to use task-based testing with paper-prototypes, carefully recording uttered thoughts and actions of the test users. Compared to the first part, this part of the test had a more general scope and concentrated on the understanding of the concepts, not on single usability issues on the interface.

The test was conducted with the same user group as described in section 3.13.1. They were seated in front of a whiteboard, on which the paper mock-ups were posted using magnet stickers. The users got a paper remote control to convey the idea of using a TV. The tablet paper prototype was handed to the user to be put on the lap.

4.2. Test Flow

4.2.1. Task Design

The test contained three tasks. For all tasks, the initial screen was the main menu of the settings application paper prototype.

1. Log in to the TV with your user profile. If you do not have a profile yet, please create one. Finally, verify if you are really logged in. Please comment aloud on your thoughts and actions.
2. Establish a permanent connection between your tablet and the TV. Verify that your connection attempt has succeeded. Please comment aloud on your thoughts and actions.
3. Activate the device connection you have just established. Log in on the tablet and open the settings application. Please comment aloud on your thoughts and actions.

4.2.2. Prototype sketches

The paper prototype visualized the main menu interface as used in the first test and offers interface solutions for the features not yet implemented in the working prototype. The paper prototypes were sketched using mybalsamiq.com, printed and enhanced by hand.

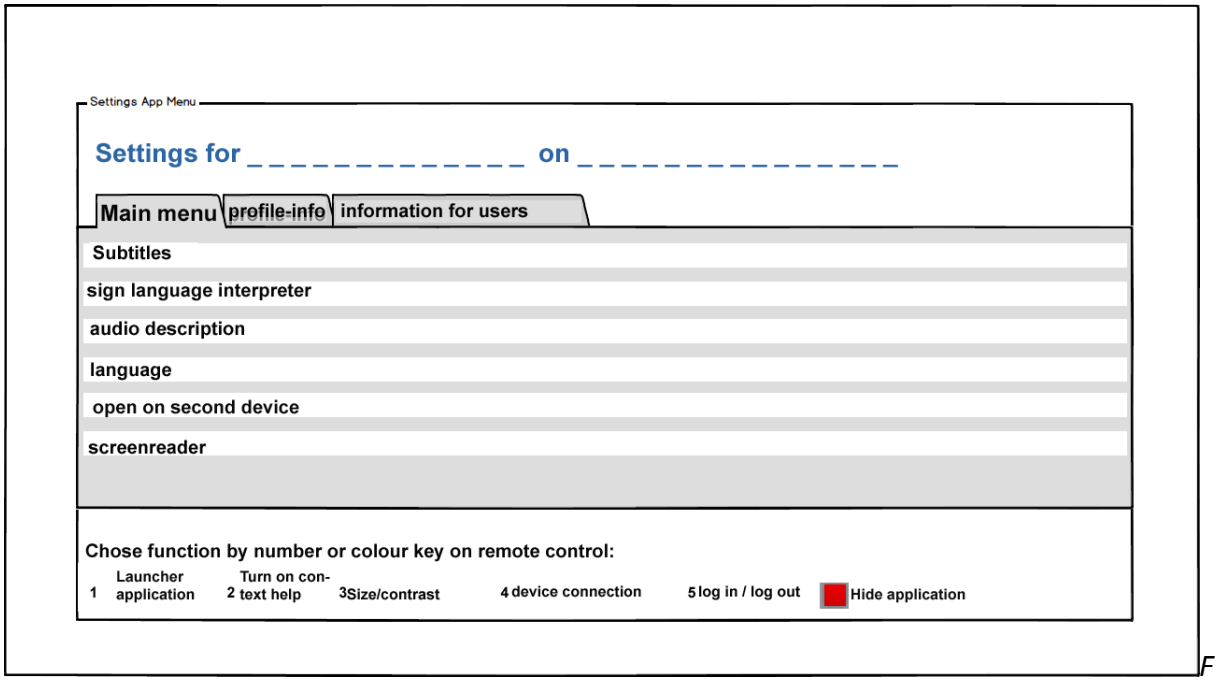


Figure 31. Paper prototype version of application main menu (see 3.3.2 for translation)

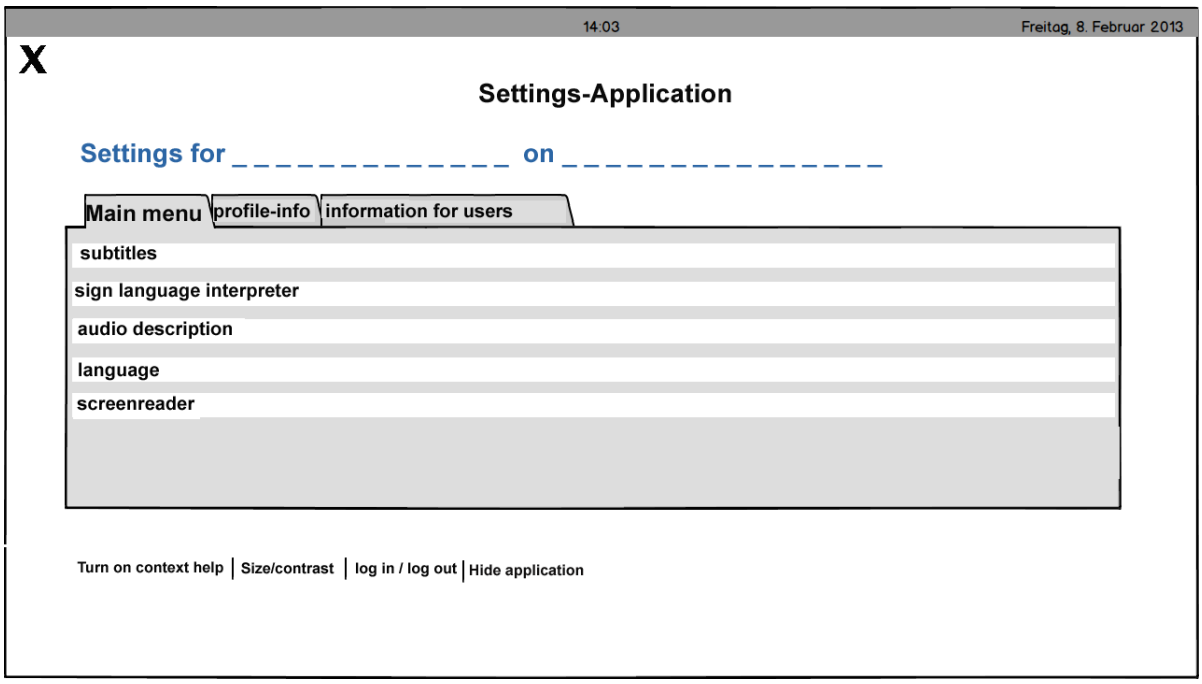


Figure 32. Main menu, tablet version

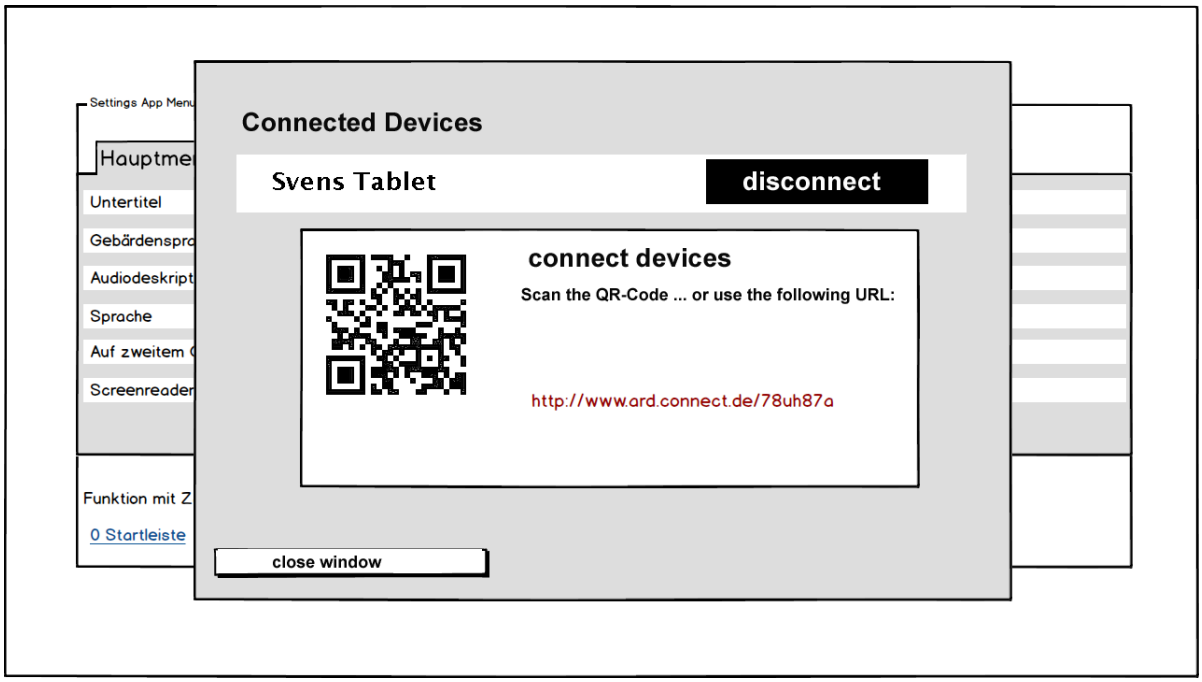


Figure 33. Dialogue for device connection

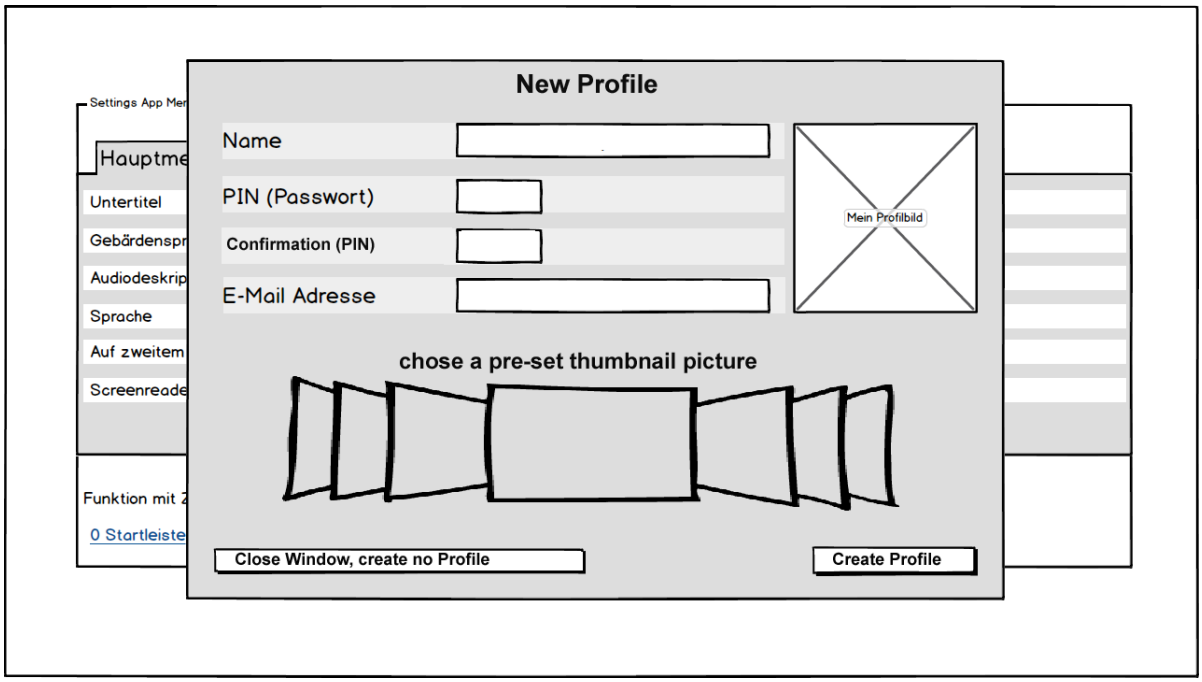


Figure 34. New user profile dialogue

4.3. Results test Nr 3: Device Coupling & Multiple Profiles

The analysis of the protocol yielded 74 issues. This report concentrates on issues that are backed by two or more users to avoid getting lost in small interface usability details irrelevant in this early conceptual stage.

It was striking that the interface seemed very complex to the test users. They frequently suggested options and explanatory texts to be added on the screen, which were already present almost literally. A reason for this may be the fact that the interface offers three menus at a time: A tab navigation on the top, the option list within the tab below and a menu of options accessible via number keys at the bottom of the screen.

Issue No.	Description and possible solution	Users	Category
1	Users state right at the beginning of the test on seeing first screen: “No, really, I do not understand that.” The new options obviously make the screen appear intimidating. **Solution**: Be less technical. Describe concepts and features from user perspective (“What can I do with that?”)	2	Concepts
3	Users choose “open on second device” first when they are told to establish a device connection. The correct option “connect devices” is listed in the bottom menu and either neglected or second choice. **Solution**: The different menus have a certain priority in perception. Similar options should be in the same menu or merged to one option at all.	7	Ambiguity

Issue No.	Description and possible solution	Users	Category
5	When the users are asked to choose a name for a newly established device connection, they enter their own name. This is often the same name as the profile name. This leads to irritations throughout the interface. **Solution**: Subsume device connections under the profile, so that only one concept is used instead of two. If the connections have to be separate from profiles, add example name to the form, like “my device xy”.	8	Concepts
9	When the connection is established and confirmed on the tablet, there are no further options on the tablet screen. Users try to either close the window (cutting the connection) or click other buttons. **Solution**: Allow for direct access of applications from the tablet, without prior transfer of the screen from within a TV application.	2	Feedback
10	The paper mock-up suggested that there was already one permanent device connection established. The users wanted to cancel those connections first before establishing their own connection. **Solution**: Obviously users assume that there can only be one connection at a time. Multi-user is an unfamiliar concept to them. This has to be explained to the users in the interface and the manual.	4	Mental Models

Issue No.	Description and possible solution	Users	Category
14	When asked to establish a device connection, the users stated confidently: “Then I must use the screenreader.” Users start guessing when no obvious choice is available. **Solution**: Use clear, self-explanatory terms in the menus. Space permitting, examples can be added to illustrate menu options.	2	Irritations
16	“What does URL mean?”, “What does link mean?”. For the wide audience of an interactive TV service, seemingly ubiquitous terms may be unknown. **Solution**: Use speaking terms like “internet address” instead of URL.	2	Wording
17	QR-Code concept is unfamiliar to users. **Solution**: Guide user step-by-step through the QR-Code process with easy explanation.	2	Concepts
19	Users asked whether they can watch TV and subtitles on the connected device. Users appear to have certain expectations toward the connected tablet, which sees the tablet more as a substitute than as a companion to the TV. **Solution**: not applicable	2	Mental Model
43	When asked to start a new profile, the users choose the “profile-info” tab, instead of “log in/out” on the bottom menu. **Solution**: Resolve ambiguity and rethink relation between bottom menu and central menu list.	6	Information architecture

Issue No.	Description and possible solution	Users	Category
46	Users fully ignore the option to set a profile picture in the new profile dialogue. This could be due to the nature of the paper prototype, but it is also likely that the users do not notice the feature because they do not expect it there (or at all). **Solution**: Personal images could be retrieved through a linked social network profile. Otherwise further research would be needed to determine whether people prefer pre-set thumbnails against no thumbnails.	3	Concepts
47	Users verify their login status at “log in / out” instead of using the foreseen “profile info”. **Solution**: The split into two menu points was obviously done based on the implementation model. This separation is not logical to the user and should be reconsidered.	2	Ambiguity
48	The users assumed that the profile was not only used in the settings application but for all applications in the launch bar. **Solution**: This is actually how the profile could work in the future. Yet the question is whether it then still would make sense to have the user log in in the “settings” application.	2	Information architecture

Issue No.	Description and possible solution	Users	Category
49	Users want to erase existing profiles first before creating their own profile. **Solution**: Like in issue 10 the users seem to be unfamiliar with the idea that there can be more than one profile at once. This means that this new concept has to be introduced and explained thoroughly in order to be understood and accepted.	4	Mental Model
64	Although the user is logged in on TV, the connected tablet shows "unknown" as user status. The user is irritated, because she thought she already was logged in. Irritation is increased by issue 5 – name confusion between profile and connection name. **Solution**: Re-design the logic of logging in and access rights from second device.	2	Consistence

4.4. Post-Test

A short post-evaluation test was designed to estimate the overall attractiveness of the new features. It uses the net promoter score (Reichheld 2006, [7]) and forces the users to choose the feature most and least important to them.

4.4.1. Net Promoter Score

The net promoter score (NPS) (Reichheld 2006, [7]) indicates the likeliness of user acceptance and commercial success of a product or service by measuring the declared willingness of the test users to promote it to others. The NPS is measured by the following scale: "How likely is it that you will recommend this application to a friend or colleague?" (0 = unlikely; 10 = highly likely)

The NPS defines those answering 9 or 10 as promoter, those answering 0 to 6 as critics and those with 7 and 8 as indifferents. The NPS is calculated as percentage of promoters minus percentage of critics.

The questionnaire identifies 2 promoters (with values 9 and 10) and two critics (with values 1 and 5). 6 are counted as indifferents (value 8). This yields a **NPS of exactly 0**. Counting only technical lead users, the NPS would be at +20, counting accessibility multiplicators only the NPS would be at about +16. These values differ much more than in the previous tests documented in D2.3.1, where the NPS was 100. A plausible explanation would be that in the first test the application concept was both very inconcrete and new to the users, which lead to an initial excitement. Now, with a working application prototype, the users can give a much more differentiated feedback.

	HBB-NEXT prototype
All users	0
Technical lead users	+20
Accessibility multiplicators	+16

It should be noted, that we do not have reliable benchmarks yet for these scores. Instead the scores assessed here shall work as a benchmark for any upcoming tests. Interpreted by themselves, the NPS here indicate that the features of the prototype are valued by the more tech-savvy users, but that on the average users are not fully convinced of the additional benefit of the new and more feature-rich application.

4.4.2. Most important feature

We asked the users which of the features demonstrated in both tests they would like to see most in a regular service. Given the target group this question is biased, as sign language and subtitles obviously cover more basic needs than profile and tablet connection.

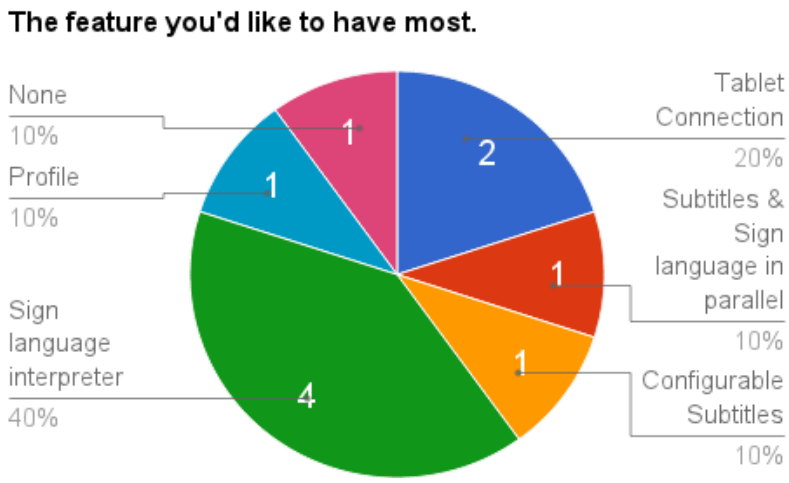


Figure 35 Most important feature as named by the participants

4.4.3. Least important feature

We asked the test users about the feature they could spare first from those introduced in the two tests. The synchronized tablet connection was chosen by half of the user (2 of them own a tablet, 3 of them do not). A reason might be that in the tests no actual use of the synchronized tablet was demonstrated, but only the connection process. The users were thus left with their own assumptions on the use cases. Two hard-of-hearing users said they would spare the sign language interpreter, as they did not speak sign language.

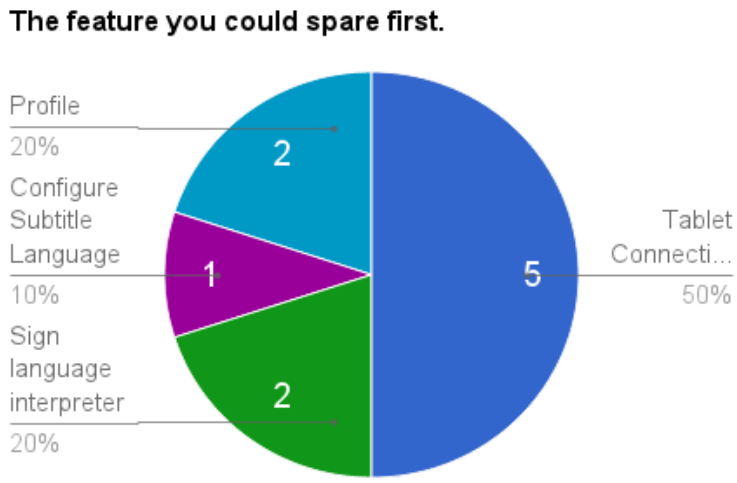


Figure 36 Least important feature as named by participants

4.4.4. Understanding of profile concept

At the end of the test we asked all users to explain what they understood by “a profile”. Only three out of ten testers were able to give a concise explanation. The others had vague ideas; three of them appeared not to be familiar with the concept of profiles from any other context.

4.5. List of Recommended Solutions

The list of recommended solutions takes into account the issues mentioned in the test results. Furthermore, it sums up the broad picture of a number of further issues, which cannot be discussed meaningfully here in detail each. The solutions are grouped into categories derived from a simple grounded theory categorization effort on the issue table.

Resolve Ambiguity

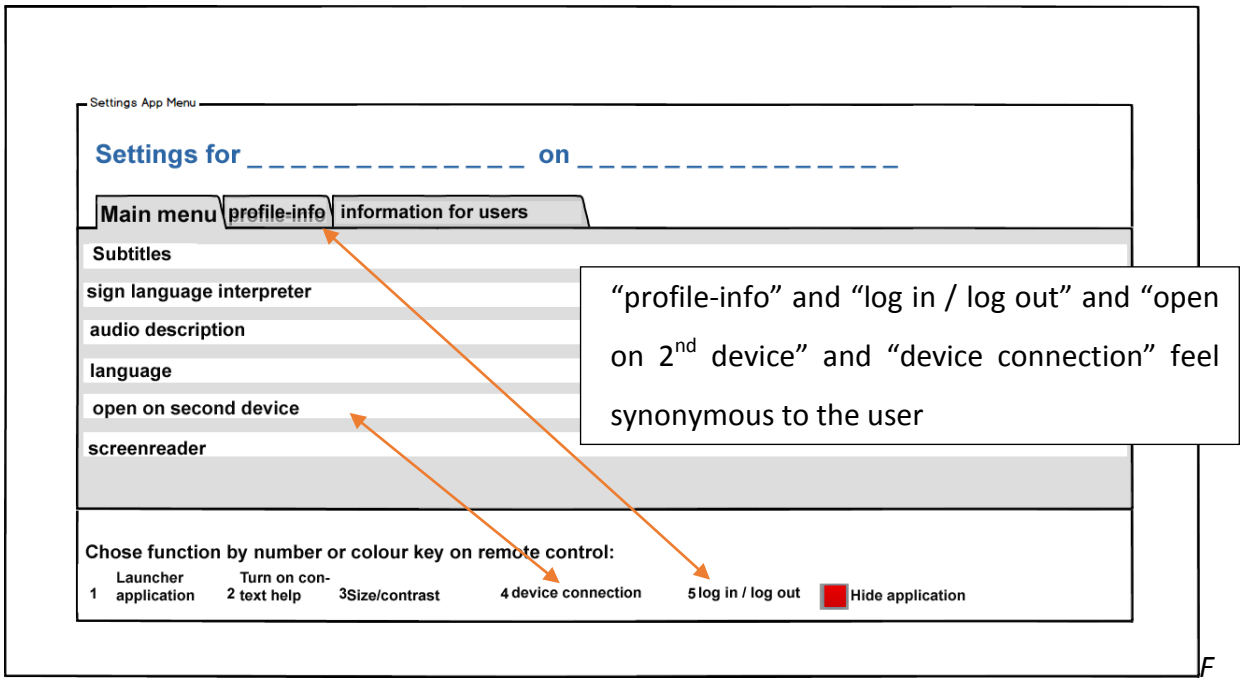
The interface offers a number of features that are ambiguous to the user:

- Profile Information vs. Logging in/out
- Open on second device vs. Device Connections

For any related task the majority of test users would try first the options that can be found directly in the menu. The options that are implemented in the framework are always second choice and considered after some searching only, as they are separated from the rest of the interface at the bottom of the screen.

The implementation logic of having a separate framework for system-wide profile handling should not be imposed on user interface level.

Furthermore, a lot of users even mix up the connection of a second device with logging in with their own profile. For example, eight out of ten users named their user profile and their tablet connection exactly alike. It should be considered to merge the two concepts closer.



igure 37. Ambiguous features on the interface

Explain Multi-user concepts

Users assume implicitly that only one user can be logged in at a time, even if more than one profile is available on the TV. If they want to connect a new device, they are tempted to remove the currently connected one. If they want to log in, they want to log out current users first. This shows that multi-user concepts on TV and tablet are fairly unknown at least in the elderly tester group we had. It has to be introduced with care and good explanations.

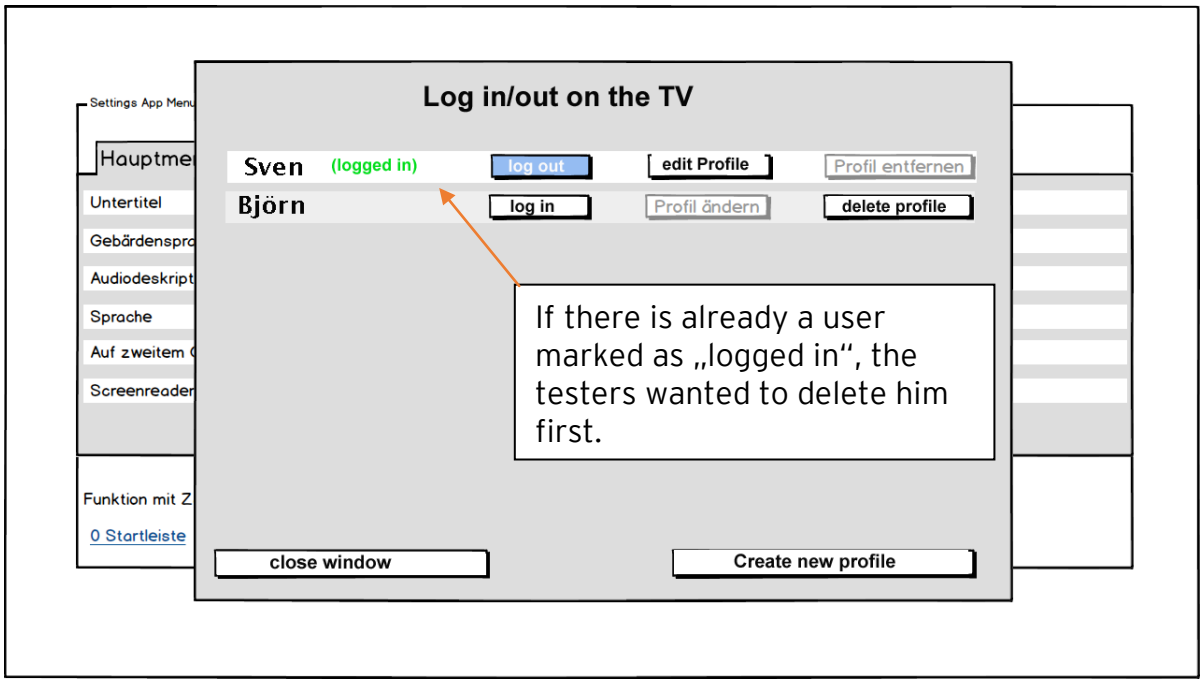


Figure 38. Log in / log out dialogue

Apply consistent behaviour across devices

When users are logged in and connect the tablet to use an app on second screen, they are puzzled that they have to log in again on this device. Also it is not clear, which settings (Group profile or individual profile) should be changed when editing on the second device. The relation between logging in and device connection should be simplified and aligned to consistency between both devices. The logic of a Master-Slave connection between TV and Tablet was not consistently implemented in our tests.

Re-think the Profile Concept

The profile concept as such was unfamiliar to about a half of the test users and even after explanations they were barely able to describe it in their own words. The other users described the profile as a way to store settings so that they would not have to change them. However, no user expressed a direct desire for this feature. Some were the only user of their TV set anyway, while others did not have a problem to agree on common settings with their partner. There should be a re-assessment of the core user needs the profile concept can solve. The concept should be simplified and its benefits should be explained to the user more clearly.

5. Conclusion

The tests show that the HBB-NEXT applications are in a usable state, i.e. it can be expected that users can use and reach their goals with them without major problems. This is reflected by the low problem rates in the quantitative evaluations of the tests. Yet both *efficiency* and *joy of use* have to be improved in order to create marketable and competitive applications. We see that a number of people still need considerable time to work through the interface. From both KU Leuven and RBB tests we find that the applications still feel too technical.

The tests regarding multi-user features and device connection show that these are new concepts, which we have to introduce to the user thoroughly. Those users who did understand the concepts by and large saw possible merits, so the risk of imposing something totally undesirable is low. Then again, the problems to explain the concepts at all to a part of the users should be taken as a reminder of a close reconsideration of user needs and applied interaction concepts.

5.1. Most important findings

KU Leuven and RBB identified common themes across all user evaluations:

- A more deliberate **terminology** should be used in all interfaces. As TV is widely spread, Smart TV applications are likely to face a very diverse audience. As opposed to the PC, the TV has always been considered a very simple device. We can assume that the users have less tolerance for tech language on TV and we should reconsider the use of seemingly familiar, but non-speaking terms like “URL” and terms describing rather technical acts than human actions like “authenticate”.
- Some **design patterns** known from the web can be applied to Smart TV interfaces and if so, they should be implemented in the way familiar to the user. However the different ergonomics and usage context of TV applications also demands for the identification of new, TV-specific design patterns. Basic visual design rules regarding contrast, font sizes and functional grouping have to be adapted to the requirements of the Smart TV interface.

In general, user interface for TV can show less information and UI elements because everything needs to be larger due to the larger viewing distance compared to PC.

- Appropriate **feedback** of the current system status is an established principle for usable interfaces. It improves the user's ability to explore and understand the interface on his own. For any operation lasting longer than just a few seconds, the application should show information about the progress to increase the trust in the application and avoid the user wondering whether the application was still running.
- **Registration procedures** to create a profile should be kept brief, with as few information asked as possible. It may be possible to edit more information later optionally, but most users want to get started right away. It is preferable to ask for information in the context where and when it is actually needed.
- **Design concepts that mirror** the underlying technical architecture are likely to seem logical to the implementer, but not to the user. This affects both terminology and separation of functions that belong together from a user perspective.

5.2. Directions for future work

In the third phase of user evaluations we aim to test the application in actual in-home settings as much as possible in order to address the real context-of-use. As the majority of usability flaws have likely been discovered by now, the methodology will shift towards quantitative evaluations on efficiency as well as the assessment of the overall user experience, including aspects such as look and joy of use. Qualitative assessments will provide us with more meaning concerning the quantitative results.

6. References

[1] Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, *7*(1), 57–78. doi:10.1080/10447319509526110

[2] Nielsen, J. (1994). Usability inspection methods. In *Conference Companion on Human Factors in Computing Systems* (pp. 413–414). New York, NY, USA: ACM. doi:10.1145/259963.260531

[3] Spencer, D. (2009). *Card Sorting: Designing Usable Categories*. Rosenfeld Media.

[4] Story, M. F. & Mueller, J.L. & Mace, R. L. (1998). The Universal Design File: Designing for People of All Ages and Abilities. In: Design Research and Methods Journal, Vol 1, No 1 (2011).

<http://design-dev.ncsu.edu/openjournal/index.php/redlab/article/view/102>

[5] Rubin, J. & Chisnell, D. (2011). Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests. John Wiley & Sons. pp. 261

[6] Goldstein, E. Bruce (2010). Sensation and Perception, Eighth Edition. Wadsworth Cengage Learning.

[7] Reichheld, Frederick F. (2003). The one number you need to grow. In: Harvard Business Review, 12/2003, p.47-54.

[8] Schenk, Michael (2007). Medienwirkungsforschung. Mohr Siebeck Tübingen.

[9] Hippel, Eric von (1986). Lead Users: A Source of Novel Product Concepts. In: Management Sciences, Vol 32, No 7 (1986).