COMPETITIVENESS AND INNOVATION FRAMEWORK PROGRAMME

ICT Policy Support Programme (ICT PSP)

ICT PSP call identifier:	**ICT PSP/2007/1**
ICT PSP Theme/ Objective identifier:	**3.0: Emerging Access Services**
Project acronym:	**DTV4All**
Project full title:	**Digital Television For All**
Grant Agreement no.:	**224994**
Contract duration:	**01.07.2008 – 31.12.2010**
Deliverable no.:	**D3.5**
Deliverable title:	**2nd Phase Emerging Access Service Demonstrators**
Nature of deliverable:	**Report**
Dissemination level	**Public**
Due date:	**31.08.2010**
Actual delivery date:	**06.09.2010**

Document title and history: **D3.5 – 2nd phase Emerging Access Service Demonstrators**			
Version no:	*Date:*	*Changes*	*Contributor*
001	18-01-2010	First draft distributed for discussion	IRT, UAB
002	30-03-2010	Contribution added	RAI
003	25-05-2010	Contribution added	TV Catalonia
004	08-06-2010	Contribution added	IRT
005	10-06-2010	Editorial rewrite	Brunel
005	22-06-2010	Contribution added	TV Catalonia
006	28-06-2010	Minor revisions	Brunel
007	19-07-2010	Contribution added	RBB
008	27-07-2010	Contribution added	UAB
008	28-07-2010	Contributions added, Revisions	RBB
009	29-07-2010	Checked and forwarded to Brunel	IRT
010	19-08-2010	Re-counted and -calculated numbers in test results	RBB
011	27-08-2010	Editorial rewrite	Brunel
012	5-09-2010	Rewrite and corrections	UAB

Document authors / organization:	**IRT et al**
Partners contributed:	**All**
Availability:	**Drafts**
Circulation:	**Partners**

Table of Contents

1. Introduction

This deliverable covers the laboratory tests of the so called phase II demonstrators carried out in the DTV4All project from the start of the project up to month 26 of the project (August 2010). The phase II demonstrators are called "Second phase Emerging Access Service Demonstrators" and encompass demonstrators for Video Signing (Signer on a TV screen and/or mobile phone), Clean Audio delivered on an additional sound channel, Reduced playback speed with the help of hard disk storage, Enhanced Audio Description via the Web and last but not least Enhanced Text Services, provided via the Web.

The current document elaborates the technological approach to realising each of the demonstrators as well as the respective test set-ups and methodological approaches. Test results of the phase II demonstrators are documented in this deliverable.

Further results which are not ready for presentation in this document are to be dealt in the successor document D3.6 called "Final Report on Expert User Tests of Emerging Access Services" which is scheduled for project month 28. The final deliverable named "Recommendations for Future Access Services, Devices and Platforms" (D3.7) will be considerably based on this deliverable. D3.7 is scheduled for project month 30 which is the last month of the project.

2. Executive Summary

This document focuses on interfaces for achieving barrier-free Digital Television that are subject to user testing by DTV4All. These interfaces are also called “emerging demonstrators” if they fulfil certain criteria. These selection criteria are:

- Represents the state of the art in respect of the technologies used
- Provides services which address needs of people with an impairment
- Deliver services that are not yet on the market at a pan European level but are expected to become available across Europe in the near future

At the time this report was compiled a complex selection process had already taken place to identify the emerging services that would be subject to user testing. This process is described in detail in DTV4All deliverables 3.1 - 3.4. The DTV4All project makes a clear distinction between mature and emerging access services. In contrast to mature access services the so called emerging access services are dependent upon the availability of new platforms and devices. The phase II demonstrators have been selected as representative of the latest technical developments for optimising access to digital television.

Chapter 3 gives a quick overview of the results of all the laboratory tests carried out on the phase II demonstrators up to July 2010. Chapter 4 contains details of the tests carried out with a demonstrator for Video Signing. Chapter 5 gives details of the tests held with a demonstrator for Clean Audio. Chapter 6 gives details of a demonstrator for Reduced Playback Speed. Chapter 7 contains details of the tests of a demonstrator for Enhanced Audio Description. Chapter 8 gives details of a demonstrator of an Enhanced Text Service. Chapter 9 describes tests with a German language demonstrator of text to speech applications.

Chapter 10 contains additional material on the various tests. This includes firstly detailed results of the tests with a Demonstrator for Text to Speech Applications in the German language. Like the other laboratory tests performed by RBB, this test was quite complex as it encompassed tasks in the second step of each individual test session and finally evaluations and direct questions as the third part. However, distinct from the other tests carried out by RBB, the idea of these tests was not to test a novel service or device developed in DTV4All. Instead, the tests were used to gather user feedback on a device already on the German market that would be applicable to other European markets for which equivalent devices are not yet available. The so called Ammec German language

text to speech device was used in the tests to derive generic, in the sense of is European level, recommendations on text to speech applications for sight impaired people. Only the generic conclusions were translated into English, the detailed German test results are contained in the Chapter 10.

Chapter 10 also contains the questionnaire for the Reduced Playback Speed tests, a modified questionnaire for the RAI tests of signing applications as well as a detailed matrix showing the composition of RBB's user group for the Enhanced Text Service tests.

3. Overview of the results

3.1 Results of the Tests with demonstrators for video signing

The tests were carried out in the subjective assessments room at the RAI Research Centre and Innovation Technology laboratory in Turin (Italy). To further improve the investigation of the Video Signing Service, the questionnaire has been slightly modified compared to the previous version reported in an appendix to deliverable D3.4. The DTV4All prototype platform was used for tests in Italian sign language. This solution allows the users to activate or disable the interpreter depending on their needs. The users can also control the position and size of the interpreter window.

It was reported that the presence of an interpreter for signing on the screen is very useful for deaf people who are able to understand the sign language.

The preferred position for the interpreter window is at the bottom on the right hand side. Some users want to have control over the interpreter window position. This should be possible using a specific button which is easily identifiable on the decoder remote controller.

Users' testing with small screens (14" and smaller) prefer to have the signer occupying half the size of the TV screen or more. Users' testing with larger screens (bigger than 14") want to have the signer smaller, down to 30% of the size of the TV screen.

A service that offers translation in sign language on a mobile device did not receive positive user feedback when it aimed to substitute the interpreter window on the TV screen. It seems to be tiring to watch an interpreter for sign language on a small hand held device like a mobile phone. There is a minimum size the interpreter has to appear on a display. When the display is too small the interpreter becomes hard to understand. A standard mobile phone screen seems to have too low a resolution for this application. There is one exception, watching a film in a cinema with a sign language interpreter on the screen of a mobile phone would be accepted by the signing deaf; the authors' of this document note that this kind of application is not covered by DTV4All, because this is a cinema application and not a TV application.

It was evident that young deaf people quickly mastered the controller because they are accustomed to digital technologies, while elder people needed more training and attention.

3.2 Results of the Tests with a Demonstrator for Clean Audio

As elaborated in Deliverable D3.4, three Clean Audio user tests were carried out with the German Language Demonstrator jointly by RBB and IRT. Two pre-tests (in April 2009 and September 2009) involving a limited number (three and five respectively) of hard of hearing testers were carried out to gain an understanding of how much attenuation of background noise and music should be done. The main test incorporating the pre-test findings took place in November and December 2009. RBB provided test material from its archive that was processed by IRT for all the tests. Like the two pre-tests, the main test was also done via a DVD that was provided to testers.

The main test often showed remarkable improvements in audibility for the test participants. The results, however, seemed to depend a lot on the type of video material that was chosen for applying clean audio. Acceptance was best when only limited processing was done, i.e. removal of restrained ambient noise or music resulting in nearly no impairment in the audio quality of the dialogue. The samples of clean-audio where a lot of ambient noise (cheering, applause) had been removed were less well received by the testers.

The testers' comments gathered by UAB can be summarised as: Clean Audio is a very good solution. However, the testers weren't aware of the difficulties (technical and financial) of producing audiovisual material with this option but all the testers thought investing in such techniques was worthwhile, particularly if clean audio could be generated automatically. All the testers thanked the team who had provided the demonstrator and produced the test materials and wanted to know when they would be able to have a clean audio service at home.

3.3 Results of the Tests with a Demonstrator for Reduced Playback Speed

Test results have clearly shown that reducing playout speed can have a positive effect on both visual processing and comprehension.

In terms of visual processing, eye-tracking analysis leads us to conclude that a reduction in playout speed can give viewers more time to not only read subtitles but also to pay more attention to the image. However, excessive playout speed reduction could be counterproductive. For the 100% playout speed versions participants' gaze could disperse due to subtitles perceived to be excessively quick while in the 70% playout speed versions the same phenomenon could take place if the subtitles are perceived to be excessively slow.

There is a clear uniformity within all the speed reduction percentages in terms of detailed comprehension. On the one hand, as expected all the reduced playout speeds gave better

comprehension results than the original playout speed. However, on the other hand, a greater reduction in playout speed does not always mean better comprehension. If the viewer perceives the further reduction in playout speed to be excessive it can be counter productive.

The conclusion of the tests is that a modest reduction in playout speed can have a very positive effect for some people who have reading or cognitive difficulties.

3.4 Results of the Tests with a Demonstrator for Enhanced Audio Description

These results are not contained in this document. At the time this deliverable was compiled the tests were not yet completed. This document contains a general description of the tests only.

3.5 Results of the Tests with a Demonstrator of an Enhanced Text Service

An HbbTV-based text service for the German nationwide Channel 1 of ARD was launched at IFA 2009. Through the cooperation of IRT and RBB a barrierfree version of this was service conceived, prototyped, and finally tested in December 2009. The idea was to create an inclusive prototype of an innovative text service for digital TV that uses the potential of new technologies to promote barrierfree access for sight impaired people. The new features included zoom capability, a number of additional colour variants to be chosen, and a still rudimentary audio component based on text-to-speech technology. The user group was constituted with the help of the local user associations in Berlin and Brandenburg, the test concept and questionnaire were also supported by advisors from these associations. The user group was representative of the most important sight impairments (see section 10.4 of this document).

The new barrier-free equipped HbbTV-ARDText was very well received by its visually impaired testers. Some testers even expressed a great deal of enthusiasm for the enhanced text service. Many of the testers found it to be a real improvement compared to traditional teletext. Testers found the customisable font size enlargement and colour/contrast settings especially helpful. They liked the "computer-like handling" and thought the menu structure was, on the whole, straight forward.

The user trial did expose a few "weak points" to be considered if this text service is to be implemented. The users were extremely engaged and constructive. They made many suggestions for improvement. These included many points that were not obvious to sighted individuals, and therefore weren't part of our initial concept of a barrier-free version.

Section 8.5 offers a detailed résumé of the task accomplishment and free commentary sections of the test. Each feature and functionality of the Enhanced Text Service is treated in detail including

suggestions for improvement from the testers' perspective. These are expected to be of generic use when thinking about recommendations for future novel text services.

3.6 Tests with a Demonstrator for Text to Speech Applications

Tests with the German language Ammec-device, a multi media device equipped with a TV receiver card, either for DVB-S, T, or Cable, which provides audio interfaces, were carried out at RBB in June and July 2009 with sight impaired testers. The envisaged laboratory test focused not on validating the Ammec device as such but on validating the concept of "Audio User Interfaces" (AUI) for TV sets / digital receivers provided with the help of a text-to-speech engine (TTS). The Ammec was used for the test because at the moment it is the only TTS-based set top box on the German market. The laboratory tests had a general and a specific aim:

General aim:

- To determine the *general acceptability* of such a service: Are spoken or audio interfaces generally welcomed by the target group, do "they make sense to the users" in terms of achieving improved access to digital television?

Specific aim:

- To find out how such a service should be designed in order to meet the requirements of the target users mainly in terms of the scope of its functionalities but also in terms of basic navigation issues.

The test results show very clearly that a TTS-based device (Set-Top-Box) providing a wide range of functionalities is very much desired by the target group. All the testers would use it and consider it very important. All the tested functionality areas were deemed 'very important' or 'important' by the testers. These include, apart from the ability to choose TV-channels, the Electronic Programme Guide and Radio-related functionalities, as well the capability to record and cut TV and radio programmes, accessing read-out Teletext services, and a little bit less importantly, playing CDs and DVDs.

The summary of the results listed in chapter 9 presents detailed results based on users' accomplishment of tasks, including free comments and observations, and on direct questions and comments after having accomplished the tasks. The results are structured into basic considerations on operational concepts for audio interfaces, on what is desired by the testers in terms of the direct functionality buttons a remote control should offer, and finally on the most prominent areas of interest that were identified by the testers. The latter include a number of functionalities, basic conceptual issues and issues related to remote controls for such services.

4. Tests with a Demonstrator for Video Signing

Partners involved: Brunel, IRT, RAI

By Andrea del Principe

4.1 Introduction

DTV4All demonstrated in 2009 a hybrid broadcast/broadband approach to delivering a signing service for the signing deaf. For a broadcast television programme a video of a sign language interpreter for the programming is transmitted via a separate broadband Internet connection and shown on a television screen synchronously with programme that is being broadcast.

4.2 Situation according to the workplan

A first prototype based on the SAVANT project prototype platform was available in 2008 and was improved on and then duplicated in 2009 by IRT and Brunel. Towards the end of 2009 a sample of this prototype was transferred to RAI which will undertook laboratory tests and user tests in December 2009.

4.3 RAI Savant based Video Signing Test details and results analysis

4.3.1 Introduction

Nowadays sign language interpretation services are offered for a restricted range of programmes, typically on news or specially made programmes. People dislike the idea of a picture partly obscured by a signer for the deaf, for this reason signed programmes are usually scheduled at unsocial hours.

The DTV4All prototype platform, based on hybrid broadcast/broadband technology, can be used to deliver the deaf signing access service as a closed (elective) service on digital TV channels. This means that the interpreter can be activated or disabled by the user.

The window with the sign language interpreter in it can also be moved and resized, allowing the user to decide the position and size of the interpreter window.

New technological developments offer new opportunities and solutions that could be applied to improve human activities and various aspects of life. The role of the research world includes analyzing possible implementations. After a phase of study and know-how acquirement, the researchers through a process of the free exchange of ideas and creative thinking identify new fields and contexts in which the new technology can be applied. The output of the creative activity is a set

of ideas. When this process is applied to access services, it is necessary to carefully identify which of the many technically feasible ideas are destined to be accepted by and useful to the end user. Also the functional requirements of the new services can be defined only after studying the needs and preferences of the users. Video signing for television programmes is a consolidated service. Since in this case DTV4All is offering a new technological solution aimed at improving an existing access service, offering new features and functionalities, the emerging service must be carefully tested, both in comparison to the traditional one and with respect to the new features it offers.

4.3.2 Test general description

The DTV4All prototype platform is used to deliver a new closed (elective) deaf signing access service on digital TV channels. This solution allows the user to activate or disable the interpreter depending on their needs. The user can also control the position and size of the interpreter window.

The tests are designed to explore two important issues.

1. Response of the users to the new service:

   The aim of this test session is to understand if the service is generally accepted and to obtain an overall evaluation of it. The test questions investigate various aspects like the "level of satisfaction", which features are considered more important etc. For example, the "level of satisfaction" is a parameter based on a range of possible answers from "not useful" to "very useful".

2. Functional requirements:

   This test session sought to determine the functional requirements of the emerging access service, with the aim of defining how the actual service should be designed to meet the users' requirements. The questions were designed to determine the best settings for a wide range of features such as the best position of the interpreter window on the screen, its preferred size, if the size should be automatically calculated by the system on the basis of screen size and viewing distance etc.

The tests were followed by an investigation into the possibility of providing the sign language interpreter on a mobile device instead of on a television screen. The study of technology required to do this is not part of the DVT4All project but the test session for DTV4All was considered a good opportunity to gather some preliminary information on the viability of such a service. The aim was simply to find out if a translation into sign language presented on a standard hand-held device, like a mobile phone or a smart phone, would be accepted by the target group.

4.3.3 Test set-up and implementation

The tests were carried out in the subjective assessments room at the RAI Research Centre and Innovation Technology laboratory in Turin (Italy).

The room was equipped with three 16/9 TV monitors of different sizes: 14", 28" and 46". Using different size screens allows for the tests to determine if the screen size influences the testers' choice of interpreter window position and size.

When prompted by the test questions, the testers' used a remote controller to change the interpreter window activation setting, position, and size.

4.3.4 Questionnaire

The test of the emerging access service was based on a questionnaire with sixteen questions.

In order to gather results that could be readily analysed, the questions posed were multiple choice questions. For example:

Would you find it useful to freely set the position of the interpreter window on the TV screen?

☐ It would be very useful

☐ It would be useful

☐ It would be slightly useful

☐ It would not be useful

The questionnaire was printed on paper and the testers made their response by marking their preferred answer with a cross. A free text field was included in the questionnaire to collect comments, notes and suggestions.

Two kinds of questions were presented to the testers:

- Questions with an associated video sample
    - A video sample makes easier the comprehension of the question. It can be demonstrated that people with a limited technological knowhow find it difficult to imagine a new application only by reading a text description. The above example question was accompanied by a video showing the interpreter window in different areas of the screen.
- Interactive questions

- The tester is asked to set the interpreter window size or position according to his or her preference using the remote controller.

4.3.4.1 Preparing and training the users

An exhaustive explanation was considered fundamental to providing the testers with all the information they needed to execute the tests. Since the first test session it had become clear that only after fully understanding all the questions were the testers able to give meaningful answers. As in this test we had to deal with signing deaf users, all the explanations were translated by a Sign Language interpreter, who in turn translated the deaf participants' questions into Italian so that the RAI technical staff present could answer them. The test sessions were supervised by specialized personnel.

The information about how to use the remote controller to set the size of the interpreter window and position it was given in Italian Sign Language. It was evident that young deaf people quickly mastered the controller, being accustomed to digital technologies, while older users needed more training and attention.

4.3.4.2 Test session details

The operator started the test session manually following all the processes set down beforehand and collected the completed questionnaires.

A complete session required an hour and half to gather feedback from four signing deaf individuals. The adopted procedure was the result of previous experience of subjective assessments being applied to this particular test. For each session, the participants were invited to sit in the assessment room where RAI specialists provided them all the necessary explanations concerning the use of the remote controller, registration, the purposes of the test, and the methodology that would be used to consolidate the testers feedback and obtain results from it. This explanation and training activity involving the four signing deaf people required about thirty minutes to complete. After that, three of them left the room and the test was undertaken the first person. Each individual test took about fifteen minutes to complete.

A test session was planned in the morning and two in the afternoon.

4.3.4.3 Participant profile

The test was undertaken by signing deaf people (deaf persons who use signing as their main language).

There were thirty-eight participants in the test. The data collected is considered enough to outline trends on a statistical basis. To allow for a comprehensive analysis it was decided to invite a heterogeneous group of individuals to participate in the test. The test group of users was composed of people with different age ranges:

- Range A: under 25
- Range B: from 25 to 50
- Range C: from 50 to 70

All participants were accustomed to the traditional Italian sign language service on TV programmes.

Among the testers there were different levels of skill in computer and mobile devices technology but they were all at least able to use the basic functions of a remote controller.

4.3.4.4 Analysis of the results

The report on this test points out various aspects, first of all the global results for each question are presented, then the tester feedback is used to identify those needs of the user that are based on their age range.

On the questionnaire the tester is requested to provide personal information such as their gender, age, and family situation etc. This additional data can be used for further analysis, making it possible to define different tester profiles, for example on the basis of gender, education level, etc.

This will potentially allow default parameters to be defined for a user's television receiver. During initial setup the receiver could ask the user to input their age range, gender etc., in order to identify a corresponding stored profile and consequently set the correct viewing preferences.

The results obtained are as follows:

Question 1

Translation: While you are watching TV, do you find it useful to have a sign language interpreter on the screen?

1. It would be very useful
2. It would be useful
3. It would be slightly useful
4. It would not be useful

Question 1	Global		Range A <25		Range B 25 to 50		Range C >50	
Response 1	26	69 %	13	72 %	8	67 %	5	62 %
Response 2	10	26 %	5	28 %	4	33 %	1	13 %
Response 3	2	5 %	0	0 %	0	0 %	2	25 %
Response 4	0	0 %	0	0 %	0	0 %	0	0 %

Table 1: Responses to Question 1

Comment: Almost all the people involved in the test gave a positive answer to question one. The sign language interpreter on the screen is considered useful or very useful. Table 1 shows that the demand for a sign language interpreter is homogeneous among the people involved in the test sessions in all the age ranges considered.

Conclusions: The presence of a sign language interpreter on the screen is very useful for Italian signing deaf people.

Question 2

Translation: Which is your favourite solution regarding the position of the interpreter window on the screen?

1. bottom - right
2. bottom - left
3. top - right
4. top – left

Question 2	Global		Range A		Range B		Range C	
Response 1	28	74 %	14	77 %	8	66 %	6	75 %
Response 2	7	18 %	3	11 %	2	17 %	2	25 %
Response 3	3	8 %	1	6 %	2	17 %	0	0 %
Response 4	0	0 %	0	0 %	0	0 %	0	0 %

Table 2: Response to question 2

Comment: Almost all the people involved in the test preferred the standard bottom – right position for the interpreter window. The table shows clearly that the bottom – left position was the next most preferred position. Only a small percentage of users (8 %) indicated a preference for the top – right position. Users commented that the preferred interpreter location depends on the main programme content. For example, sometimes the results of the sport matches appear in the bottom right corner.

Conclusions: The preferred position for the interpreter window is bottom – right. For the video signing decoder/receiver, this position should be adopted as the default position.

Question 3

Translation: Would you like to freely set the position of the interpreter window on the TV screen?

1. It would be very useful
2. It would be useful
3. It would be slightly useful
4. It would not be useful

Question 3	Global		Range A		Range B		Range C	
Response 1	31	81 %	16	89 %	11	92 %	4	49 %
Response 2	6	16 %	2	11 %	1	8 %	3	38 %
Response 3	0	0 %	0	0 %	0	0 %	0	0 %
Response 4	1	3 %	0	0 %	0	0 %	1	13 %

Table 3: Responses to question 3

Comment: Almost all the people involved in the test gave a positive answer to question three; users found the ability to control the position of the sign language interpreter on the screen very useful or useful. The opinion is homogeneous among the participants regardless of their age. Users commented that some programmes give scores, subtitles or useful information in a corner of the screen. The feature under analysis is considered very important because it can improve user comprehension and viewing experience.

Conclusions: People involved in the test found very useful the investigated feature. Suggestion: The user should be able to set the interpreter window position operating on a specific button easily detectable on the decoder remote controller. A four directions button was suggested by the people interviewed on occasion of the test sessions.

Question 4

Translation: Please, set the interpreter window size on the 14" screen in front of you using the remote controller.

1. preferred size: 10 %
2. preferred size: 20 %
3. preferred size: 30 %
4. preferred size: 40 %
5. preferred size: 50 %
6. preferred size: 60 %
7. preferred size: 70 %
8. preferred size: 80 %
9. preferred size: 90 %
10. preferred size: 100 %

Question 4	Global		Range A		Range B		Range C	
Response 1	0	0 %	0	0 %	0	0 %	0	0 %
Response 2	0	0 %	0	0 %	0	0 %	0	0 %
Response 3	2	5 %	2	11 %	0	0 %	0	0 %
Response 4	12	32 %	10	56 %	1	8 %	1	13 %
Response 5	20	52 %	6	33 %	9	75 %	5	62 %
Response 6	4	11 %	0	0 %	2	17 %	2	25 %
Response 7	0	0 %	0	0 %	0	0 %	0	0 %
Response 8	0	0 %	0	0 %	0	0 %	0	0 %
Response 9	0	0 %	0	0 %	0	0 %	0	0 %
Response 10	0	0 %	0	0 %	0	0 %	0	0 %

Table 4: Responses to question 4

Comment: In order to evaluate and compare results obtained using different screen dimensions, a viewing distance corresponding to 8 H was set, i.e. 8 times the height of the screen (BBC R&D White Paper – "*Results of a survey on television viewing distance"* - N.E. Tanton). This viewing distance takes into account the viewing habits of the most Italian users.

On a 14" screen the chosen interpreter size was 50 % or 40 % (responses 5 and 6), this means that the sign language window covered about one-half of the screen.

Furthermore, considering the results on the basis of the different age ranges, the preferred window size increased with age. As shown in Table 5, in age range C 25% of the users preferred a window covering 60% of the main programme.

In the free text field present for each question, many testers suggested adopting a sort of chroma-key solution for the interpreter. In this way at a given interpreter size, removing the background of the sign language window would preserve a bigger proportion of the main programme.

Conclusions: Most of the testers involved in the test with a 14" TV set the preferred interpreter window dimensions corresponding to 50% or 40% of the TV screen.

Question 5

Translation: Please set the interpreter window size on the 28" screen in front of you using the remote controller.

1. preferred size: 10 %
2. preferred size: 20 %
3. preferred size: 30 %
4. preferred size: 40 %
5. preferred size: 50 %
6. preferred size: 60 %
7. preferred size: 70 %
8. preferred size: 80 %
9. preferred size: 90 %
10. preferred size: 100 %

Question 5	Global		Range A		Range B		Range C	
Response 1	0	0 %	0	0 %	0	0 %	0	0 %
Response 2	0	0 %	0	0 %	0	0 %	0	0 %
Response 3	2	5 %	2	11 %	0	0 %	0	0 %
Response 4	20	53 %	13	72 %	5	42 %	2	25 %
Response 5	11	29 %	2	11 %	4	33 %	5	62 %
Response 6	5	13 %	1	6 %	3	25 %	1	13 %
Response 7	0	0 %	0	0 %	0	0 %	0	0 %
Response 8	0	0 %	0	0 %	0	0 %	0	0 %
Response 9	0	0 %	0	0 %	0	0 %	0	0 %
Response 10	0	0 %	0	0 %	0	0 %	0	0 %

Table 5: Responses to question 5

Testers involved in the test were able to see and understand all of the sign language interpretation when the interpreter window covered about 40% of the whole screen. Taking into account the responses to the previous question, it is possible to outline a general trend: the preferred percentage of the area of the screen covered by the interpreter window decreases when the screen size increases.

Considering the results on the basis of age range, the preferred window size increases with age. As shown in Table 5, moving from age range B to age range C there is a significant tendency to prefer a larger interpreter window size. In range C most of the users (62 %) preferred an interpreter window that covered 50 % of the screen. Older users tend to prefer a wider interpreter window, probably due to a physiological reduction in viewing ability with age.

In the free text field present for each question, many testers suggested adopting a sort of chroma-key solution for the interpreter.

Conclusions: Most of the testers involved in the test with a 28" screen TV in age ranges A and B preferred an interpreter window with dimensions corresponding to the 40% of the whole screen, for age range C the preferred dimensions corresponded to 50% of the screen.

Question 6

Translation: Please set the interpreter window size on the 46" screen in front of you using the remote controller.

1. preferred size: 10 %
2. preferred size: 20 %

3. preferred size: 30 %
4. preferred size: 40 %
5. preferred size: 50 %
6. preferred size: 60 %
7. preferred size: 70 %
8. preferred size: 80 %
9. preferred size: 90 %
10. preferred size: 100 %

Question 6	Global		Range A		Range B		Range C	
Response 1	0	0 %	0	0 %	0	0 %	0	0 %
Response 2	4	11 %	3	17 %	1	8 %	0	0 %
Response 3	27	71 %	13	72 %	8	67 %	6	74 %
Response 4	5	13 %	2	11 %	2	17 %	1	13 %
Response 5	2	5 %	0	0 %	1	8 %	1	13 %
Response 6	0	0 %	0	0 %	0	0 %	0	0 %
Response 7	0	0 %	0	0 %	0	0 %	0	0 %
Response 8	0	0 %	0	0 %	0	0 %	0	0 %
Response 9	0	0 %	0	0 %	0	0 %	0	0 %
Response 10	0	0 %	0	0 %	0	0 %	0	0 %

Table 6: Responses to question 6

Users involved in the test were able to see and understand all of the sign language interpretation when the interpreter window covered about 30% of the whole screen; Taking into account the responses to the previous question, it is possible to outline a general trend: the preferred percentage of the area of the screen covered by the interpreter window decreases when the screen size increases.

Considering the results on the basis of age range, the preferred window size increases with age. As shown in Table 6, in age ranges B and C the preferred interpreter window size was bigger than in range A. Older users tend to prefer a wider interpreter window, probably due to a physiological reduction in viewing ability with age.

In the free text field present for each question, many testers suggested adopting a sort of chroma-key solution for the interpreter.

Conclusions: The most part of the people involved in the test with a 46" TV preferred an interpreter window with dimensions corresponding to the 30% of the whole screen.

Question 7

Translation: Would you like to see a sign language interpreter on your mobile phone interpreting a TV programme you are watching?

1. I prefer to see the sign language interpreter on the TV screen
2. I would like to see it on a mobile phone
3. It would like both solutions

Question 7	Global		Range A		Range B		Range C	
Response 1	25	66 %	11	61 %	8	66 %	6	75 %
Response 2	3	8 %	1	6 %	2	17 %	0	0 %
Response 3	10	26 %	6	33 %	2	17 %	2	25 %

Table 7: Responses to question 7

Comment: For the most part the testers preferred to see the sign language interpreter on the TV screen.

Some testers expressed a preference for response 3 - "I would prefer both solutions". This response requires further analysis. It is difficult to determine if there is a real need for a mobile TV signing service or if this response was given just to preserve the possibility of such a service being made available.

In the free text field, many people reported serious difficulty watching a TV programme and simultaneously following the translation by looking at the mobile phone.

Conclusions: A service that offers the sign language interpretation on a mobile device does not seem to receive positive feedback when it is aimed at substituting an interpreter window on the TV screen.

Question 8

Translation: Would you like to see sign language interpretation on your mobile phone while you are watching a movie at the cinema?

1. It would be very useful
2. It would be useful
3. It would be slightly useful
4. It would not be useful

Question 8	Global		Range A		Range B		Range C	
Response 1	20	52 %	12	67 %	6	50 %	2	25 %
Response 2	10	26 %	4	22 %	4	33 %	2	25 %
Response 3	4	11 %	2	11 %	2	17 %	0	0 %
Response 4	4	11 %	0	0 %	0	0 %	4	50 %

Table 8: Responses to question 8

Comment: Almost all the users involved in the test gave a positive answer to question eight. A service providing a sign Language interpreter on a mobile device is considered useful or very useful while enjoying a movie at the cinema. Table 8 shows that the demand for a sign language interpreter a mobile phone screen aimed to be used at the cinema is homogeneously accepted among the people of age ranges A and B. Being a quite innovative service, people in age range C are divided about it. One half considers it useful or very useful, while the other half seems not to be interested in it.

In the free text field, some users commented that the service could be provided directly by the cinema, as happens in some movie theatres in USA where a small display with subtitles is located on the back of the seat in front of the viewer.

Conclusions: A service that offers a sign language interpreter on a mobile device aimed to be used at the cinema is considered very useful because it would allow a deaf user to enjoy a movie with other deaf or hearing people.

Question 9

Translation: Please make a comparison, which do you think is more tiring, watching the sign language interpretation on a mobile phone screen or watching it on a TV screen?

1. It is much more tiring
2. It is slightly more tiring
3. There is not any difference

4. It is less tiring

Question 9	Global		Range A		Range B		Range C	
Response 1	23	60 %	14	78 %	6	50 %	3	38 %
Response 2	12	32 %	4	22 %	4	33 %	4	43 %
Response 3	3	8 %	0	0 %	2	17 %	1	13 %
Response 4	0	0 %	0	0 %	0	0 %	0	0 %

Table 9: Responses to question 9

Most of the users involved in the test considered it more tiring to view the sign language interpretation on a mobile device than viewing it on TV.

In fact many users reported serious difficulty in following the interpreter on a mobile phone.

Conclusions: A service that offers sign language interpretation on a mobile device does not receive positive feedback when it is aimed at substituting the interpreter window on the TV screen.

Question 10

Translation: Gender:

1. Female
2. Male

Question 10	Global		Range A		Range B		Range C	
Response 1	23	61 %	11	31 %	6	50 %	6	75 %
Response 2	15	39 %	7	39 %	6	50 %	2	25 %

Table 10: Responses to question 10

Question 11

Translation: Education:

1. Elementary School
2. Middle School
3. High School

4. University

Question 11	Global		Range A		Range B		Range C	
Response 1	16	42 %	7	39 %	4	33 %	5	62 %
Response 2	13	34 %	6	33 %	5	42 %	2	25 %
Response 3	6	16 %	3	17 %	2	17 %	1	13 %
Response 4	3	8 %	2	11 %	1	8 %	0	0 %

Table 11: Responses to question 11

Question 12

Translation: When did you become deaf?

1. Born deaf
2. Under 3 years
3. Over 3 years

Question 12	Global		Range A		Range B		Range C	
Response 1	29	76 %	16	83 %	9	75 %	4	49 %
Response 2	6	16 %	2	11 %	1	8 %	3	38 %
Response 3	3	8 %	0	0 %	2	17 %	1	13 %

Table 12: Reponses to question 12

Question 13

Translation: What is your degree of deafness?

1. Hard of hearing
2. Deaf

Question 13	Global		Range A		Range B		Range C	
Response 1	33	87 %	15	83 %	10	83 %	8	100 %
Response 2	5	13 %	3	17 %	2	17 %	0	0 %

Table 13: Responses to question 13

Question 14

Translation: In your family where either of your parents deaf?

1. Both parents
2. One parent
3. Neither of them

Question 1	Global		Range A		Range B		Range C	
Response 1	11	60 %	4	22 %	4	33 %	3	38 %
Response 2	4	29 %	2	11 %	1	8 %	1	13 %
Response 3	23	11 %	12	67 %	7	59 %	4	49 %

Table 14: Responses to question 14

To further improve the investigation of the Video Signing Service, the questionnaire has been slightly modified compared to the previous version reported in an appendix to deliverable D3.4. Please find the final updated version used during the tests reported here in Chapter 10 of this document.

5. Tests with a Demonstrator for Clean Audio

By Sven Glaser, Oliver Pidancet, Bettina Heidkamp, Martin Link, Pilar Orero

Partners involved: IRT, RBB, TVC, UAB

5.1 Introduction

Often there is a need to clean up dialogue for listeners with hearing impairments to provide Clean Audio (or Clean Sound) as the original sound is mixed with ambient noise or music. "Clean Audio" means the "cleaned up" dialogue is delivered without any other significant audio components in an extra audio channel provided by the broadcaster. The DVB standard allows several audio tracks to be assigned to an MPEG video/audio stream. DTV4All will focus on the delivery of Clean Audio content over traditional broadcast distribution platforms like satellite, cable and on-air networks.

Clean Audio is an emerging access service that is expected to be a first step in improving intelligibility for large user groups of hearing impaired people despite the extreme heterogeneity of hearing impairments. For instance when several persons in a room are talking simultaneously a listener with normal hearing can catch up easily and follow the discussions; this is known as the "cocktail party effect". Some people are not able to understand a single person in such an environment. In this project IRT tries to satisfy the needs of most of hearing impaired persons.

Clean Audio cannot be used as a substitute for a hearing aid. Clean Audio is a service that could be helpful to many people with a slight or medium hearing impairment.

An implementation of Clean Audio by DTV4All based on a proprietary IRT solution was demonstrated at the European Ministerial e-Inclusion Conference 2008 in Vienna.

Successive laboratory tests were undertaken by RBB and UAB in late 2009 and in spring 2010. The results of RBB's two pre-tests are documented in Deliverable 3.4. They can be summarised as follows. A majority of users found remarkable improvements in audibility through clean-audio applications, but the rest did not show utter satisfaction with the whole set of test material. Most wide acceptance was in the case when only smart processing was done, i.e. removal of only restrained ambient noise or music resulting in nearly no impairment in the audio quality of the dialogue. The samples of clean-audio where a lot of ambient noise (cheering, applause) had been removed were less well received by the testers.

The main RBB test incorporating the pre-test findings took place late November, early December 2009. The results of RBBs main test and of UAB's tests will be presented in the following sections.

5.2 Technological Approach

The focus of the evaluations was audio source material with no separate dialogue track, i.e. there must be a technical aid to process the audio material for suppression of noise and music accompanying the dialogue. The target was the evaluation of how existing techniques can be applied rather than inventing new strategies for the development of new acoustic audio processors.

Suppressing noise is not the same as eliminating it. In general audio content with soft music or noisy floors could be "made clean" much more effectively than dialogue mixed with loud sound elements. In the latter cases the dialogue suffers from loss of speech quality with only limited success in suppressing the noise parts.

Hence two kinds of strategies could be examined for the hard-of-hearing audience. First of all a reduction only of soft music or noise components could be offered to potential users for assessment, as the pre-test showed that the audio processing produces nearly noise-free speech. On the other hand content with loud ambient noise could be processed using different levels of attenuation to evaluate the trade-off between noise reduction and resulting speech quality.

Even without anticipating different kinds of hearing impairments a unitary result was not to be expected. Probably some listeners would be bored when they hear no sound whilst no one is talking, some others might prefer a bigger loss in audio quality when noise suppression is better, maybe only low frequencies are impairing, and so on.

5.3 Test Set-Up and Methodology for the Catalan Language test

The TVC test material/footage was received by IRT at the beginning of 2009.

This material in the Catalan language was processed by IRT for UAB until end of October 2009. The processing of this footage was done in the same manner as that used in the 2nd pre-test in the German language. Tests were carried out with several test persons in November 2009 at UAB.

5.4 Test Set-Up and Methodology for the main test carried out at RBB

5.4.1 Introduction

As elaborated in Deliverable D3.4 all in all three Clean Audio user tests were carried out with the German Language Demonstrator jointly by RBB and IRT. Two pre-tests (April 2009 and September 2009) involving a limited number (three and five respectively) of hard of hearing testers were carried out to establish how far the attenuation should go. The main test incorporating the pre-test findings took place in November/December 2009. RBB provided test material from its archive that was processed by IRT for all the tests.

5.4.2 Test Set-Up and Methodology

Like the two pre-tests, the main test was also done via a DVD that was provided to testers. Those (eight) testers who did not have a DVD player at home were provided with one by RBB. The DVD was sent out to the testers on 20th November together with a detailed questionnaire including an introduction to DTV4All, Clean Audio technology and, of course, detailed instructions that led them through answering the questions item by item. The test persons were asked not to change hearing aids and head sets but keep their "usual living room procedures" in order to obtain realistic results.

The DVD was structured into nine sections ("items"). Each section presented a different video sample in different variations (processed at different levels, unprocessed). Users had to evaluate the audibility of these variations by comparing "clean" with the unprocessed audio output using two different approaches depending on the respective item:

a) Comparison of two different Clean Audio versions one after another in a single sample video ("item")

b) Comparison of the unprocessed audio with two or three versions of differently filtered audio in a single sample video ("item").

Sections ("items") also incorporated tasks (rating with school marks) about how well a video sample version could be heard and the recognition of differences between several versions of Clean Audio.

5.4.3 User Profiles

The user group consisted of 18 persons, 8 women and 10 men, who had been participating in RBB's DVB-subtitling test from February 2009 to October 2009. Five of these had already taken part in one of the two pre-tests. Once the DVB-subtitle test has been concluded we approached all hard of hearing testers and were very pleased how many testers agreed to take part (69%) in yet another test. As the table below shows, the test group was quite heterogeneous as to their degrees of hearing impairment and as to which assistive devices they used.

5.5 Evaluation and Results

The evaluation was done as follows. Each section comprises at least two questions and a marking task (marking audio/listening quality with school marks). For each item the questions and tasks were described in detail and finally quantitative and qualitative results were extracted.

For the statistical evaluation of the marks "very good" (1) and "good" (2) were grouped as "good", satisfactory (3) and sufficient (4) as "average" and "bad (5) and very bad (6) as "bad". For each section the Clean Audio/filtering parameters or levels were listed and from this a conclusion was drawn for each section ("item"). At the end, an overall conclusion was drawn.

Additionally the different degrees of impairment outcome were grouped into 0-30, 50-70 and 80-90 in order to consider this criterion in the evaluation of the results.

All the test persons wear hearing aids. One test person uses a so called ITE device, 14 test persons used BTE devices. One person used a cochlea implant ("bionic ear") – see explanation below:

In-the-ear (ITE) aids have their working parts in the earmould so the whole aid fits into the ear.

Behind-the-ear (BTE) hearing aids usually have an earmould, which sits inside the ear. The hearing aid rests behind the ear and a plastic tube connects it to the earmould.

A **cochlear implant** (CI) is a surgically implanted electronic device that provides a sense of sound to a person who is profoundly deaf or severely hard of hearing. The cochlear implant is often referred to as a **bionic ear**.

The following matrix shows the grade of hearing impairments of the 18 test persons:

			Tester																	
Grade of hearing impairment	**Loss of hearing**	**ear**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
Low	20-40%	right																		
		left	x																	
Average	40-60%	right			x															
		left			x															
High	60-80%	right				x	x	x	x	x	x									
		left					x	x	x	x	x	x								
close to deafness	80-95%	right										x	x		x	x	x	x	x	
		left		x		x							x	x	x	x	x	x	x	
Deafness	100%	right												x						x
		left																		x
degree of impairment outcome			0	15	30	50	50	50	50	50	50	50	70	70	70	70	70	70	80	90
Type of hearing aid			**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
ITE Device		right							x											
		left							x											
BTE Device		right			X	x	x	x		x	x	x	x		x	x	x	x	x	
		left			X	x	x	x		x	x	x	x	x	x	x		x	x	
Cochlea Implant		right																		x
		left																		x
additional assistive devices			**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
induction loop						x		x		x		x	x	x	x		x		x	x
head phone									x		x			x		x		x		

Summary degree of impairment outcome

degree	*number*
0	1
15	1
30	1
50	7
70	6
80	1
90	1

Item 1:

This sample consists of two identical video segments, the first featuring "normal", unfiltered audio, the second a Clean Audio version.

Processing method: The video sample can be characterized as having a normal level of ambient noise. It was chosen to investigate if attenuation of ambient noise can enhance the audibility of dialogue. Here, ambient noise contains a hospital's background noise including soft rumbling, rustling and footsteps. Most of the appearing audio spectral components were not part of dialogue. Because of that the bigger part of ambient noise could be attenuated in the first pass of the Cedar processing. The ambient music was a simple tune and its single-tone components could be attenuated (but not completely removed) by lowering the level of the respective fundamental oscillations.

Question 1: Did you notice any difference between the normal and the Clean Audio version?

The Clean Audio parameters used with this sample were noticeable for the majority of the test persons: differences were noticed by 14 out of 18 testers, while four found no variation in the two versions.

Question 2: If yes, could you hear the dialogue in the Clean Audio version better?

Twelve out of those 14 testers replying found that they could hear the dialogue in the Clean Audio version better. The remaining two testers who were able to differentiate could not hear any improvement. One of them noted that audio worsens with Clean Audio, one that there was no improvement. These facts complement the finding of Question 1 – the Clean Audio parameters used with the Item 1 sample work effectively for a majority of the test persons: twelve out of 18.

Summarizing:

For 66.7% of all testers the Clean Audio version brought an improvement in audibility. For 5.6% each it either made no difference or there was even deterioration with Clean Audio. 22.2% of all testers were not able to identify the Clean Audio version.

Marking: Please rate with a school mark how well you could hear the Clean Audio version (1=very good…6=very bad).

50.0% of all testers evaluated the Clean Audio (CA) version as "good to very good", 50.0% as "average". These marks support the conclusions of question 1 and 2: half of the testers could agree on the Clean Audio parameters used with this sample.

	Answers	in %
Good to very good	7	**50.0%**
Average	7	**50.0%**
Bad	0	**0.0%**
	14	**100.0%**

note

Good to very good: 1-2

average: 3-4

bad: 5-6

Degree of impairment:

66.7% of the people having a degree of impairment of 0-30 were able to differentiate and were judging the Clean Audio as "good to very good".

Within the degree of impairment of 50 to 70 46.2% of the people judged the Clean Audio as "average", while 30.8% found it "good to very good". The remaining 23.1% were not able to differentiate between the two versions.

All testers having a degree of impairment of 80-90 recognized differences between the versions and judged the Clean Audio also as "average". This leads to conclusion that the Clean Audio level is quite satisfying across the different levels of hearing impairment while it is judged better by those who can hear better (0-30).

Comments of testers:

Testers commented the Clean Audio version of this first item very differently. Some judged it as well working, that background sounds were faded back, and speech was felt to be much clearer. Others stated problems in differentiating the two versions and therefore in judging. A few, however, were simply not able to distinguish between Clean Audio and the normal version. For some the dialogue did not benefit from the filtering, Clean Audio was judged as being too artificial. In a few cases testers stated that they even heard worse in the Clean Audio version, the speech was felt to be too much in the foreground.

Conclusion item 1:

As the musical was left audible in the Clean Audio version, the difference in the sound did not become apparent to the testers until the speakers began talking. The piano tones during the last 2 sentences were additionally filtered with a narrow band filter using the processing tool *Izotope* as the results with only cedar-processing were not satisfactory. Speech quality suffered from this difficult passage because of similar spectral components in the voices and in the background music. Probably for some listeners only the sequence with the piano tones was their focus for scoring, when the voices became artificial.

Item 2:

As for Item 1, this sample consists of an unfiltered audio segment and a Clean Audio version. In this sample only female voices appear which have a different frequency responses compared to male voices.

Processing method: In this video sample the dialogue is mixed with music and chirping birds, with the help of *Izotope* the latter could be eliminated nearly completely. An additional reworking using the Cedar processor was necessary. For the effective attenuation of the music components band-elimination filters were used. This resulted in slightly tainted speech. The aim of the Item 2 sample processing was to investigate how tainted speech and an effective attenuation of ambient components can be traded off.

Question 1: Do you notice any difference between the normal and the Clean Audio version?

Differences were noticed by 11 out of 18 testers, slightly more than half of the testers, while seven found no variation between the two versions. Compared to Item 1 where 14 testers out of 18 heard a difference this number is slightly lower. This could be an indication of the sample being slightly more complex, having especially higher frequencies of ambient noise including clattering dishes and singing birds which might influence speech overtones. On the other hand it is interesting to note that item 8 features a quite similar example but gave much better results.

Question 2: If yes, could you hear the dialogue in the Clean Audio version better?

Only five out of the 11 testers who had noticed a difference in the two versions found that they could hear the dialogue in the Clean Audio version better. Six of the 11 could not hear the Clean Audio version better, four of them judged both versions equally good and two the clean audio version worse than the unfiltered version.

Summarizing:

For 27.8% of the 18 testers the Clean Audio version brought an improvement in audibility, for 22.2% it made no difference while 11.1% even noticed deterioration in audibility with Clean Audio. 38.9% of the 18 testers were not able to identify the Clean Audio version.

Results of marking:

The task here was: Please rate with a school mark how well you could hear the Clean Audio version (1=very good…6=very bad).

	Answers	in %
Good to very good	6	**50.0%**
Average	5	**41.7%**
Bad	1	**8.3%**
	12	**100.0%**

50.0% of all testers rated the Clean Audio version as "good to very good". In turn, only 8.3% marked the filtered audio as "bad". With altogether 91.7% there is a broad majority of people actually stating that the Clean Audio version is satisfying with more than half of them having been able to hear it good to very good.

These numbers are much more positive than the results elaborated above for question 2 where less than 28% of all 18 testers had stated that the Clean Audio version was better for them. This could be due to the fact that all testers had to mark the Clean Audio version not only those who had perceived a difference with the Clean Audio version as in question 2.

Degree of impairment:

Only 33.3% of the testers having a degree of impairment of 0-30 were able to differentiate the Clean Audio from the unfiltered audio. They judged the Clean Audio as "good to very good".

Within the degree of impairment of 50-70, 21% of the testers judged the Clean Audio as "good to very good", another 36% found it "average" while only 7% judged it as "bad". The remaining 36% were not able to differentiate the Clean Audio from the unfiltered audio.

Only half of the testers having a degree of impairment of 80-90 were able to differentiate the two versions and they judged the Clean Audio version as "good to very good".

These results seem to show that, in contrast to the previous example, the tested type of filtering/Clean Audio is in this case especially suited for people who have a grade of hearing impairment of 50-70 but not for those whose hearing is better or worse.

Comments of testers: Again, testers' comments were very varied. For instance one person found the unfiltered audio generally better, one person stated that the audibility of the dialogue got worse with Clean Audio while another stated that with Clean Audio "the speech is in foreground but sounds tinny". Although several testers found the two versions hard to distinguish, most of the commenting persons noted that the music was softened and the dialogue appeared cleaner with Clean Audio.

Conclusion item 2:

The conflicting and inconsistent comments underline the diverse results of the marking process and of questions 1 and 2. A lot of subjectivity seems to come in.

Like the processing on item 1, the narrow-band portions of the music during speech were first swept by the *Izotope* notch filters. When no one was speaking, the noise of the dishes and the music were kept audible. The whole dialogue came with soft music strokes and ambient sounds containing lower frequencies. Both were removed quite well, probably the missing low frequency noise gave the "tinny" impression.

Item 3:

Again, a video segment was repeated, first with unfiltered audio, then in a Clean Audio version. In this sample the voice originates from a radio broadcast featuring in the video, so that there is no lip reading support.

Processing method: At the end of this video sample loud applause appears and starts interfering with the voice of the radio host. That's why the signal-to-noise ratio is much lower than in Item 1 or 2. The aim is to evaluate if a moderate attenuation of lively ambient noise can improve the audibility of dialogue for members of the target group.

Question 1: Do you notice any difference between the normal and the Clean Audio version?

Differences were observed by 13 out of the 18 testers which meant that only five persons found no variation in the two versions.

Question 2: If yes, could you hear the announcer in the Clean Audio version better?

Only three out of the 13 testers who did notice a difference in the two versions found the dialogue in the Clean Audio version more audible. Of the remaining ten, five could hear both versions equally well and five thought the audibility of the clean audio version worse than the unfiltered version.

Summarizing:

Only for 16.6% of the 18 testers the Clean Audio version brought an improvement in audibility. The people who benefit from the Clean Audio version are thus clearly outnumbered by those who found no improvement. For 27.8% of the testers Clean Audio made no difference in terms of audibility and for another 27.8% of the testers there was even deterioration in audibility with Clean Audio. A further 27.8% of the testers were not able to identify the Clean Audio version.

Results of marking:

The task here was: Please rate with a school mark how well you could hear the Clean Audio version (1=very good…6=very bad).

	Answers	in %
very good to good	3	**23.0%**
Average	5	**38.5%**
Bad	5	**38.5%**
	13	**100.0%**

Only 16.7% of the 18 testers rated the Clean Audio version as “very good to good”. Adding percentage to that of the persons who rated the Clean Audio version as “average” the percentage of the testers who found the Clean Audio version to be satisfying or acceptable is 44.4%. This is only 5.5% higher than the number of testers who were not satisfied with the Clean Audio version (marking it “bad”). The proportions of the positive responses (very good down to average) are more or less in line with the results of Question 1 and 2. However, some of the testers who were not able to differentiate the two versions may additionally have rated the Clean Audio version as “bad”.

Degree of impairment:

66% of the testers having a degree of impairment of 0-30 were able to differentiate the two versions but one half of them judged the Clean Audio as “bad” the other half as “good to very good”.

Within the degree of impairment of 50-70 only 15% of the 13 testers concerned judged the Clean Audio as “good to very good”, another 38% found it “average” and 23% judged it as “bad”. The remaining 24% were not able to differentiate the two versions.

Only 66% of the testers having a degree of impairment of 80-90 recognized differences between the two versions and half of those 66% judged the Clean Audio as “bad” the other half as “good to very good”.

Comments of testers: The comments reflect the negative tendency of the testers’ opinions described above. Both versions are described as having blurred speech by some testers, while others found the unfiltered speech clear despite the loud ambient noise. The quality of the Clean Audio version is again described as bad but a few improvements were noticed, like a softer sound, less noise in the voice from the radio and a more natural female voice.

Conclusion item 3: The results for this item seem to show clearly that the testers’ demands were not met in this sample. In this item it was not possible to completely remove the ambient noise without partly destroying the quality of the speech, i.e. some noise components remained audible. Obviously a partial removal of ambient noise results in negative impressions. The example provided was characterized as unnatural and also as no better than the unprocessed versions of items 1 and 2. Furthermore it was remarked that none of the speakers were visible whilst talking (no lip reading was possible), other samples partially show the speakers when they talk. This leads to the conclusion that problematic background noise should not be removed when no acceptable result can be achieved.

Obviously loud ambient noise in the original soundtrack decreases acceptance of the processed version which has either incompletely removed noise or poor audio quality.

Item 4:

This sample consists of three identical video segments having different audio versions: the first unfiltered, the second Clean Audio version 1 and the third Clean Audio version 2.

Processing method: Similarly to Item 1, in this sample a lot of indoor ambient noise in present, while the level of background music is relatively high. For Clean Audio version 1 an automated filtering approach was chosen (Cedar). For Clean Audio version 2 the procedure described above

was complemented by manually processing spectral components of the music. The aim of the Item 4 sample processing is to investigate how different levels of processing effort can be traded off.

Results of marking:

The task here was: How well could you hear the different versions? Please, rate with a school mark (1=very good…6=very bad).

	Answers			**in %**		
	Normal	first Clean Audio Version	second Clean Audio Version	**Normal**	**first Clean Audio Version**	**second Clean Audio Version**
very good to good	4	7	11	**22.2%**	**38.9%**	**61.1%**
average	11	9	6	**61.1%**	**50.0%**	**33.3%**
bad	3	2	1	**16.7%**	**11.1%**	**5.6%**
	18	18	18	**100.0%**	**100.0%**	**100.0%**

Taking into account only the marks “very good to good”, the second Clean Audio version is judged to be the best with 61.1 %. With 38.9 % of the votes the first Clean Audio version is judged to be quite good while the unfiltered audio only reaches 22.2 % of the votes in this comparison.

In contrast, when considering just the votes for “average”, the normal and unfiltered audio is judged best with 61.1 % of the votes.

The second Clean Audio version also got the best results in terms of the mark “bad”: While the normal and first Clean Audio versions received 16.7 % and 11.1% bad marks, the second Clean Audio version received only 5.6%.

Question: Does one of the Clean Audio versions sound exceptional pleasing or displeasing?

The testers’ judgement was quite varied: five testers found the first Clean Audio version exceptionally pleasing and six testers found the second Clean Audio version exceptionally pleasing. Two other testers, however, judged the second Clean Audio version as exceptional displeasing, which in the end shows a proportion of 5:4 in support of Clean Audio version 1.

Degree of impairment:

66% of the testers having a degree of impairment of 0-30 judged the normal audio as "average", while another 33% found it "good to very good". Both Clean Audio versions were equally judged "good". This shows that for people with a lighter hearing impairment the Clean Audio versions bring a clear improvement.

Within the degree of impairment of 50-70 54% of the testers judged Clean Audio version 2 to be the best version, with "good to very good" marks, while Clean Audio version 1 and the unprocessed audio were "average" only.

Of the testers having a degree of impairment of 80-90 33% judged Clean Audio version 2 as "good to very good", while 66% found both Clean Audio version 1 and the normal audio "average".

Comments of testers:

Once more, the various versions were described very differently. There was only one comment concerning the "normal" audio, describing it as being "muffled" and having too much bass. Clean Audio version 1 was described as sounding "tinny" with the voice being too much in the background, while others stated this version was "more understandable" and pleasing because of its "brighter sound". Clean Audio version 2 is described differently as well, comments included, "trebles are too much in foreground", it is sizzling, it has "blurred speech" but it lacks ambient noise and is therefore "more pleasing".

Conclusion item 4:

Here two different levels of processing were presented, firstly Cedar only, secondly with a preceding spectral processing, again manually done with *Izotope*.

The nearly non-impairing additional use of the *Izotope* notch filters was recognized with slightly more favourable votes.

The Cedar processing was applied excessively but the preceding filter process with *Izotope* was recognised by the testers anyway.

Item 5:

As for Item 4, here a video segment is repeated, first an unfiltered version, then two different Clean Audio versions.

Processing method: The main features here were loud and sudden indoor sounds like barking and yelling on the one hand and normal speech of varying volume levels on the other. Sudden sounds (barking) during speech and during speech pauses were attenuated with less intensity in Clean Audio version 1 and with more intensity in Clean Audio version 2.

Results of marking:

The task here was: How well could you hear the different versions? Please, rate with a school mark (1=very good…6=very bad).

	Answers			in %		
	Normal	first Clean Audio Version	second Clean Audio Version	**Normal**	**first Clean Audio Version**	**second Clean Audio Version**
very good to good	5	6	6	**27.8%**	**33.3%**	**33.3%**
average	8	7	7	**44.4%**	**38.9%**	**38.9%**
bad	5	5	5	**27.8%**	**27.8%**	**27.8%**
	18	18	18	**100.0%**	**100.0%**	**100.0%**

In contrast to the previous example, the first and the second Clean Audio versions have the same results for “very good to good”, “average” and also for “bad”. However, both are judged slightly better in the marking category “good to very good” than the normal version. The most significant marking here is 44.4% “average” for the unfiltered audio, which is 5.5% more than for each of the Clean Audio versions. Summarizing, the marking suggests that the Clean Audio versions are not perceived by the testers to have a clear advantage over the normal audio which is even judged slightly better (a lower percentages of “bad” marks).

Question: Does one or both Clean Audio versions sound exceptional pleasing or displeasing?

In contrast to the marking results, the testers’ judgement here shows a tendency against normal unfiltered audio, only one person found it pleasing and two testers judged it as being displeasing. What is in line with the marking is the fact that both Clean Audio versions are again judged

similarly: version 1 has four votes for “pleasing” and one for ”displeasing”, version 2 five for “pleasing” and two for “displeasing”.

Degree of impairment:

Within the group of people having a degree of impairment of 0-30 33% judged Clean Audio version 1 best, as “good to very good”, while version 2 and the unprocessed audio were found to be “average”.

Within the degree of impairment of 50-70 Clean Audio version 2 was judged best, having 33% “good to very good” votes, while Clean Audio version 1 got “bad” marks from 42% of the testers and the normal audio was judged to be “average” by 42% of the testers.

The testers having a degree of impairment of 80-90 judged Clean Audio version 2 best, as “average”, while version 1 and normal audio were judged to be “bad”.

Comments of testers:

The unfiltered audio was again described as being muffled and having too much bass, while another person found the ambient noise too loud but, compared to the Clean Audio versions, having a realistic sound. The comments for Clean Audio version 1 are once more differing widely, and include its volume is “too low”, changing into a brighter sound but efficiently attenuating the ambient noise. The testers found Clean Audio version 2 to have “artificial but clear” speech and that it has distorting shouts, while another person commented that he could now hear short words very well. Further comments were: “noise is nearly absent” but the volume now is “even lower”. One tester found that Clean Audio version 2 provided no enhancement in audibility at all.

Conclusion item 5:

Summarizing all the results and comments above, the advantage of Clean Audio is less convincing in this example: The marks even show a preference for Normal audio. The comments are too diverse to draw a generic conclusion. As to the grade of hearing impairment in relation to the results it seems that those with a lower grade of hearing impairment do see an improvement with Clean Audio version 1 while those with a weaker hearing capability prefer Clean Audio version 2.

This scene had several noise pulses (steps, strokes) in pauses in speech and only a rather quiet noise floor during speech (less than in items #1 and #2) had to be suppressed. Only in a short dialogue sequence was the suppression of noise, that of a dog barking, an invasive modification.

Again the problem of the trade-off between realistic presentation and attenuation with better understanding arises. The excerpt showing a dispute suffers mainly from suppressed noise whilst viewing a fight. This result affirms the basic idea that Clean Audio should not be an on/off feature, but it should be scalable by the viewer. As with item 3 some portions of the video were too difficult to make "clean" and therefore are best left unprocessed.

Item 6:

Here, in contrast to the previous items, the different audio versions (one unfiltered and three Clean Audio) are consecutively arranged within a continuous video stream and not as repetitions of an identical video segment with different audio tracks.

Processing method: The attention in Item 6 is on analysing the success of filtering very loud ambient noise. Here, in contrast to Items 4 and 5, three different Clean Audio versions were produced. Investigation is of the trade-off between deeper filtering and the loss of audible information. The latter is especially challenging because even for people without a hearing impairment the voice of the reporter at the end of the scene is hard to hear. Item 6 is not built of repeated sequences but of a continuous stream, so that this time the testers cannot remember certain speech contents, which is problematic when the hearing impression of the previous segment was better than the current. Clean Audio version 1 was done by running a *Cedar* process twice, Clean Audio version 2 was additionally passed through a broadband band-elimination filter and for Clean Audio version 3, Clean Audio version 1 was complemented with an additional pass through *Cedar* system.

Question 1: How well could you hear the dialog within each segment? Please rate the different segments with a school mark (1=very good…6=very bad).

	Answers				in %			
	Normal	first Clean Audio Version	second Clean Audio Version	third Clean Audio Version	**Normal**	**first Clean Audio Version**	**second Clean Audio Version**	**third Clean Audio Version**
very good to good	2	4	3	0	**11.1%**	**22.2%**	**16.7%**	**0.0%**
average	4	3	6	8	**22.2%**	**16.7%**	**33.3%**	**44.4%**
bad	12	11	9	10	**66.7%**	**61.1%**	**50.0%**	**55.6%**
	18	18	18	18	**100.0%**	**100.0%**	**100.0%**	**100.0%**

It is particularly noticeable that the marking "bad" shows highest values for each version. Clean Audio version 2 has the best "acceptable" (average to very good) result: 50.0%. All other versions show less than 50% acceptance.

Question: Which Clean Audio version sounds agreeable?

Mostly agreeable is Clean Audio version 1, regarding the "agreeable" attribute it has 6 votes for and only one against. With the exception of Clean Audio version 2 (3 votes for and 2 against), the other versions have a negative agreement ratio.

Degree of impairment:

Within the group of people having a degree of impairment of 0-30 the normal audio was judged as "bad", while Clean Audio versions 1, 2 and 3 were all judged "average".

Within the degree of impairment of 50-70 all four audio versions were judged as "bad" in contrast to all previous items which had more differentiated grading between the versions. Clean Audio version 3 got the least bad marks with 46% of testers judging it "average" versus 54% judging it "bad".

Those testers having a degree of impairment of 80-90 judged Clean Audio version 2 the best but only as "average", the other versions were judged as "bad".

Comments of testers:

Only one person commented the unfiltered audio, he found the music too loud. Another person found the speech of Clean Audio version 1 “good to hear” and the music noticeably quieter. For Clean Audio version 2 testers gave two contrary comments: one noticed that the speech is too quiet, the other found the “speech too accentuated and sizzling” and that it sounded artificial. The remarks regarding Clean Audio version 3 are all quite negative. Speech sounds unclear sometimes, is generally too sharp and brings unpleasant noise, while another tester found the speech softer but nevertheless harder to understand.

Conclusion item 6:

In this example the results each of the Clean Audio versions are relatively poor. However, taking into consideration all the quantitative and qualitative results, the most positive feedback goes to Clean Audio Version 2 but only for those testers with a grade of hearing impairment of 0-30.

Some portions of this sequence were not easy to listen to even with no hearing impairment, as the ambient noises, cheering and applause, were very loud. Compared to all the other samples, including the unprocessed ones, the processing results were quite poor. I.e. after processing the remaining noise floor was louder than in other unprocessed items. The quite low subjective votes (only sometimes, reflecting only small improvements in audibility) are representing this well. Item 6 had no repetitions, the movie just continued whilst toggling the different versions. This approach should give rise to different results because the testers were not able to remember as much of the dialogue when it was not repeated. The same test, but with repetitions was done in Item 7. The first "cleaning" step was preferred slightly; further degrees in "cleaning" were not preferred. During the whole excerpt the speaker was not shown (so no lip reading was possible).

Nowadays technologies for Clean Audio for very difficult and complex audio situations are limited, this is proven here. Sometimes the intention of the movie’s director is not to have all the spoken understood. This could be another limitation for applying Clean Audio. In these cases no Clean Audio process should be applied.

Item 7:

This sample consists of four identical video segments featuring different audio versions: the first unfiltered, the second Clean Audio version 1, the third Clean Audio version 2 and the fourth Clean Audio version 3. Again a lot of background noise occurs and the speakers cannot be seen.

Processing method: All samples in Item 7 were processed in the same way as those of Item 6. Similar content was used but here the same sample was shown repeatedly with the different versions.

Results of marking:

The task here was: How well could you hear the dialog within each segment? Please rate the different segments with a school mark (1=very good…6=very bad).

	Answers				in %			
	Normal	first Clean Audio Version	second Clean Audio Version	third Clean Audio Version	Normal	**first Clean Audio Version**	**second Clean Audio Version**	**third Clean Audio Version**
very good to good	3	6	4	2	16.7%	**33.3%**	**22.2%**	**11.1%**
average	4	1	6	10	22.2%	**5.6%**	**33.3%**	**55.6%**
Bad	11	11	8	6	61.1%	**61.1%**	**44.4%**	**33.3%**
	18	18	18	18	100.0%	**100.0%**	**100.0%**	**100.0%**

Just considering the marks for "very good to good", in this item Clean Audio Version 1 receives the best marks. However, with 33% this tendency is not very strong. Combining "very good to good" and "average" as an "acceptable result", Clean Audio version 3 takes the lead with 66.7% acceptance and Clean Audio Version 2 comes second with 55.5% acceptance. Most "bad" marks go to the normal unfiltered version and Clean Audio Version 1 both with 61.1% of testers judging them as bad. Going by "acceptability" the statistical values show better results the more filtered the audio is. However, the first Clean Audio version has much better results in terms of good and very good votes.

Question 2: Which Clean Audio version sounds agreeable?

Mostly agreeable is Clean Audio version 1, it has five votes for and only one against. All the other versions have an even (unfiltered 2:2) or negative agreement ratio (2:3 for Clean Audio version 2 and 2:4 for Clean Audio version 3).

Degree of impairment:

Within the group of people having a degree of impairment of 0-30 Clean Audio Version 1 was judged the best but only as “average”, while all the other versions got worse “average” and “bad” marks.

Within the degree of impairment of 50-70 Clean Audio version 1 was judged the best version with 31% for “good” but 61% for “bad”, in contrast to Clean Audio version 2 with 23% for “good” but only 46% for “bad”. Another important part of this undecided Item is Clean Audio Version 3 with 54% for “average” and only 15% for “good”.

The testers having a degree of impairment of 80-90 rated this Item badly; all versions were judged as “bad”.

Comments of testers:

This time there were no comments on the normal audio. Clean Audio version 1 was evaluated as having a “good differentiation of speech and ambient noise”. Clean Audio version 2 was described as sounding “too sharp”. Clean Audio version 3 was evaluated by one tester as having a “good differentiation of speech and ambient noise” but others found the speech unclear, wavering and the ambient noise too loud. Other comments were that the speaker’s voice is too low and altogether the sound is too artificial. These comments question the relatively good “acceptable” marks in the above table.

Conclusion item 7:

The results show a slight preference for Clean Audio versions 1 and 2 (which does not mean these versions are judged as good and very good, but as average or satisfying). Looking at the grade of hearing impairment this impression is reinforced but mainly for those having a hearing impairment with grade 0-30 who found these versions “satisfying”. For those with a hearing impairment of 50 to 70, Clean Audio version 3 is the best, however, it is only judged as “Satisfying to sufficient”. Others who have a more severe hearing impairment found it unacceptable. Unfortunately Version 3 has bad values in the question on its agreeability to listen to and also the comments on it do not show much positive feedback.

In contrast to Item 6, where a continuous sequence was used, in this Item the same excerpt was used for the unprocessed and the Clean Audio versions. The processing used to produce the Clean Audio versions was the same as in Item 6. As a result this item shows that repetition of the same excerpt or a continuously played scene causes no difference in the ratings by the test users. Again, during the

whole excerpt the speaker was not shown (no lip reading was possible). Because of the same processing as in Item 6 only small variations in the ratings for items 6 and 7 could be found.

Item 8:

This sample consists of two identical video segments. The first having "normal" unfiltered audio, the second a Clean Audio version.

Processing method: The attenuation of the background music was difficult here because of its spectral similarity to the voices. Only in speech pauses was it applied with high intensity. The results from the Clean Audio pre-tests are utilized for analysing if attenuation in speech pauses can result in better or more pleasant hearing impressions.

Question: Do you notice any difference between the normal and the Clean Audio version?

Differences between the audio versions were noted by 14 out of the 18 testers while four found no difference in the two versions.

Results of marking:

The task here was: If yes, could you hear the dialogue in the Clean Audio version better? Please rate with a school mark how well you could hear the Clean Audio version (1=very good…6=very bad).

	Answers	in %
very good to good	11	**78.6%**
average	2	**14.3%**
bad	1	**7.1%**
	14	**100.0%**

The Clean Audio version of Item 8 was judged to be pretty good, only 7.1% of testers rated it as "bad". For 78.6% of the testers it was "good to very good" which means for them it was a clear improvement on normal audio.

Degree of impairment:

66% of the people having a degree of impairment of 0-30 were able to differentiate the two versions and all of those 66% judged the Clean Audio as "good to very good".

Within the degree of impairment of 50-70 54% of the testers judged the Clean Audio as “good to very good”, another 15% as ”average” and 8% judged it as “bad”. The remaining 23% were not able to differentiate between Clean Audio and normal audio.

All of the testers having a degree of impairment of 80-90 recognized differences between the versions and judged the Clean Audio as “good to very good”.

Comments of testers:

Here for the Clean Audio version one can find two contrary groups of comments: for some testers, the speech and ambient noise sound choppy and distorted. Other testers could hear the speech better and clearer and found it “well-balanced” through a “lowered ambient noise” level.

Conclusion item 8:

The positive statistical marking values for the Clean Audio version are reinforced when looking at the grades of hearing impairment. The Clean Audio seems to be an improvement for all grades of hearing impairment. However, the best effect is achieved for those with lighter hearing impairments.

This item had louder music passages than Item 2, but nearly no other noise. Dedicated, non-invasive processing (again using *Izotope*) delivered a decrease in the music level, but parts of the music were still present Again, the majority of testers stated that Clean Audio is a success for improved intelligibility of dialogue material when only small levels of ambient sound are reduced. This matches the experiences with very loud background noise when Clean Audio processing showed no improvements in audibility.

Item 9:

Similarly to Item 8 two versions of the same sequence are used, an unfiltered and a Clean Audio version.

Processing method: Similarly to the previous item, the Item 9 audio was processed in the speech pauses only. But this time the sequence contains a larger proportion of speech.

Question: Do you notice any difference between the normal and the Clean Audio version?

Differences were noticed by 16 out of 18 testers, the best result in the whole test, while only two testers found no difference in the two versions.

The task here was: If yes, could you hear the dialogue in the Clean Audio version better? Please rate with a school mark how well you could hear the Clean Audio version (1=very good…6=very bad).

	Answers	in %
very good to good	7	**43.8%**
average	7	**43.8%**
bad	2	**12.4%**
	16	**100.0%**

For 43.8% of the 16 testers who where able to hear the difference, the Clean Audio version is an improvement. Adding the "average" result one can see an acceptance of 77.8% for the entire test group of 18 persons.

Degree of impairment:

66% of the testers having a degree of impairment of 0-30 were able to differentiate between the two versions and judged the Clean Audio as "good to very good".

Within the degree of impairment of 50-70 36% of the testers judged the Clean Audio as "good to very good", another 36% found it "average" and 14% judged it as "bad". The remaining 14% were not able to differentiate between the two versions.

66% of the testers having a degree of impairment of 80-90 recognized differences between the two versions and judged the Clean Audio version as "good to very good".

Comments of testers:

Only a few comments were given here. Two testers found the Clean Audio speech too blurry, another one noted that the volume is too low in Clean Audio. The last comment was that sibilants are now "less sizzling".

Conclusion item 9:

When looking at the "good and very good" results for this item the results for the Clean Audio version are not as good as in the case of item 8. However, when looking at acceptability (very good to satisfying) they are equal. Again: the lower the grade of hearing impairment is, the better are the results.

Some reminders of the underlying music and the chirping birds remained audible during speech in the Clean Audio version, in speech pauses the ambient audio is continued silently. During speech especially the tonal components were removed using *Izotope* filters.

Technically this item matches well to items 1, 2, and 8, but the speakers' voices are softer and calmer. Probably due to this, the positive impression of the Clean Audio version is smaller than in the corresponding preceding items.

5.6 Results of the Clean Audio Test in the Catalan Language at UAB / Barcelona

Tests for Clean Audio were carried out with 10 people. Nine of them had substantial hearing-loss. One had minor hearing-loss.

All participants were educated with university degrees and their ages ranged from 50 to 82.

Their comments can be summarised as: Clean Audio is a very good solution. They weren't aware of the difficulty (technical and financial) in producing audiovisual material with this possibility, but 100% thought to be worth of investing in such technique, in particular, if it could be done automatically. All the users thanked the team who had thought of the system and produced the testing materials, and wanted to know when they would be able to enjoy it at home.

Testing conditions were with users wearing their hearing aids, since they all had problems with sound under/above a certain variable threshold. All experienced difficulties with ambient noise. They sat down in front of their TV in the conditions they watch it at home on a daily basis.

Materials used for the tests were considered acceptable, since the speed of the conversation, language and accent were those used in Catalan TV.

Video clips 1, 2, and 3 were very difficult to understand without Clean Audio even with the volume at maximum. Participants commented on the problems with the ambient noise in clip 1, which made it impossible to hear the actual words of the conversation; With Clean Audio it is very easy to understand, and volume can remain at normal setting.

The results for these three items with Clean Audio were spectacularly positive.

For video clips 4, 5, and 6 the conversation was understood but with the volume at a very high setting. The participants also had to pay special attention. With Clean Audio the reception was more relaxed, the volume of the TV could go down to a normal level, and it was a nicer experience.

Comments. Since the same clip is repeated so many times, people managed to memorise what they understood, and that allowed them to concentrate on the parts they had special difficulty with.

Even though there is no disputing the huge improvement in audibility with Clean Audio, if more tests were to be carried out it is suggested that a first test could be carried out in order to choose the level of sound, and once established, that could remain fixed in the rest of the tests.

5.7 Resume of the Clean Audio tests from a technical perspective

Though the number of test participants was limited, a quite heterogeneous group of participants as to their type of hearing impairment could be acquired for this test. From the test results we learned that some hearing impaired persons do not benefit from Clean Audio ("I missed the music…"). But if only a certain number of participants preferred the "clean" or some "cleaner" sound over the original audio, "Clean Audio" could at least be a basic principle for individual audio presentation in the receiver box. The level of the ambient noise suppression could then be scaled by the individual user selecting a mixture of the clean track and the normal audio. In speech pauses the normal audio could appear according to the individual settings. Audio scenes with difficult situations, i.e. loud ambient noise, could come with (scalable) signalization to the receiver box so that during these scenes the box would play the normal sound when the appropriate user setting is tuned to it.

Due to the nature of the available technologies for Clean Audio it has to be prepared manually and because of this it needs additional effort to prepare a broadcast including a Clean Audio track. Actually a lot of material is easy to broadcast with a Clean Audio service, as it has separate clean dialogue tracks available, especially when during audio-production a hard-of-hearing audience has been anticipated. It should not be forgotten that the test results confirm the need for Clean Audio program services. Dialogue tracks from the original recording site are of better quality than those distilled from final audio mixes like those used in the DTV4All Clean Audio listening tests.

6. Tests with a Demonstrator for Reduced Playback Speed

Partners involved: UAB, IRT

By Marta Miquel Iriarte

6.1 Background

Dyslexics and some people with cognitive impairments could benefit from receivers that allow the playing speed of the video/audio content of a program to be reduced. It is expected that the comprehension for highbrow content like scientific contributions etc. will be improved and conversations will become easier to understand. This is particularly true if pitch compensation is applied which is of high importance for understandability and acceptance.

Digital TV receivers or set-top boxes with the capability of reducing the play out speed of broadcast content including subtitles, in real time are not yet on the market. The situation is different for stored content on DVD: Many DVD players offer this possibility. The handicap of all these players is their inability to do pitch compensation. This means, the pitch of the audio will be decreased when reduced playback speed is applied. The intended purpose on DVD players is navigation and not to increase the understandability of the content. These DVD players are not applicable and cannot be used for the intention for this project.

6.2 Tests with uniformly reduced playback speed applied by UAB, content delivered by TVC

UAB will focus on uniform slowdown with pitch compensation but without time correction. This means, the overall play out time of the footage will be increased by the same factor as the slow down applied. The reason for this choice is the availability of PC based tools allowing pitch compensation together with a reduced playback speed of audio visual content. If the tests foreseen in this project are positively received the likelihood of any future implementations of reduced play out speed capabilities in Digital television receivers or set-top boxes will be increased. It is expected that the rate of deceleration cannot become more than 50% compared to the original speed without reducing the comprehensibility considerably.

The content used for the test must be such that it is hard for dyslexics and people with cognitive impairment to understand. For that reason footage was selected from Genres like talk shows, sport and science programmes.

The tests were carried out in Spanish language.

6.3 Set up of a digital mock-up of a test environment for uniform slowdown

In the first instance, IRT has determined the appropriate rate of slow down using free software. There are players in the World Wide Web able to do slow down with pitch compensation. All of them are PC based using the PC's hard disk. The players are applicable for stored content only and cannot be used for streaming content which a TV program usually is. This deficit is compensated for by the capability of doing the pitch compensation. As a consequence of the uniform slowdown the viewing time is prolonged proportionally. The benefit for people with cognitive impairments could be better understandability and comprehensibility of content. Appropriate content was provided by UAB taken from Spanish TV. Some PC based software solutions for reduced playback speed were tested by IRT and applied to UAB's footage. As a result of this process, a DVD was compiled that has been used for laboratory tests at the facilities of UAB. The results of this initial test will be given to Brunel for further enhanced tests. UAB has arranged expert tests in the Spanish language. Technical support was provided by IRT.

The next section is aimed at presenting the way in which the experiment was designed. Precise details about the participants and the material used in the study will be given as well as details as to how the experiment was conducted. Afterwards, the assumptions and premises on which the experiment was base will also be presented. Finally, the methodology used in the evaluation of the results will be explained.

6.3.1 Participants

Eight native Spanish speakers (three women and five men) participated in the experiment. All of them hold a college degree or higher. All the participants have advanced reading skills and they were all accustomed to reading subtitles, particularly when watching movies and TV-series. None of the eight participants had either hearing or visual impairments. Four of them acted as the control group (hearing Spanish speakers) the other four, in spite of being also hearing people, carried out the experiment under deaf conditions.

Thus, when the control group watched the clips the three possible inputs were active. That is to say, they received the visual input, the auditory input (the original dialogue in Spanish) and the subtitle-input in Spanish. The other four participants watched the various clips without the soundtrack in order to explore the reception of the videos when viewers cannot receive (or understand) the information, either completely or partially, based on the auditory input. In this sense, that could not

only be the case of deaf or hearing impaired people, but also of elderly people or people without a good understanding of Spanish.

Subjects were asked to watch the videos at the normal playback speed, as if they were simply watching a subtitled program on TV at home. They were asked to watch all together four pairs of videos. The participants watched the videos in an eye-tracker Tobii T60 monitor (more details are given in the section Analysis). After watching each pair of videos, they answered a questionnaire. They were informed that the experiment would probably last between 20 and 40 minutes, no time limit was set to fill out the comprehension questionnaires. Experiment sessions generally lasted 40 minutes in total.

6.3.2 Material

In order to explore the impact of different slowed-down speeds, eight clips were selected from "59 Segundos", a live TV debate that is broadcast on a weekly basis on TVE, the Spanish Official Television. As "59 Segundos" is one of the few live programs that is subtitled in real-time in Spain, the choice of this program corresponds to the aim of using real material. Slowed-down speeds could be restrictively applied to live programs of this kind, in which the visual input is not critically relevant when understanding the information. In TV debates the camera is usually focused on head-and-shoulder shots either of the TV host or of the guests/collaborators and medium range shots of the television studio set (see Figure 6.1 below).

Unlike in other TV broadcasting, such as films, TV-series or sports events, the image is not so important or indeed useful in understanding the information offered in the program. Thus, in this kind of programs both the auditory and subtitle-input become relatively more important.

The technique used in the subtitling of "59 Segundos" is respeaking. As Eugeni (forthcoming) points out:

> (…) respeaking is a new technique increasingly used by broadcasters to subtitle live programmes. From an operational point of view, the respeaker hears the original source text and orally repeats (*verbatim* respeaking) or reformulates (*non verbatim* respeaking) it, with the shortest delay possible, into a version that will be comprehensible and readable in written form. Since respeaking is mainly devoted to accessibility of audiovisual live products for deaf and hard of hearing people, the respeaker has also to edit his/her verbal output by orally or manually inserting punctuation and colour change to identify when a new speaker begins to speak. The respoken input is then transformed into subtitles by a speech-to-

text recognition software programme, which projects them onto the television screen in real time (with a 3-6 second time lag), as the programme is broadcast (Eugeni).

GUESTS/COLLABORATORS' SHOTS

MEDIUM RANGE SHOTS

Figure 6.1: Shots extracted from the material used in the experiment.

Bearing in mind that the punctuation also has to be dictated, it is clear that respeaking is not an easy task. In this sense, if the speaker, in this case the TV host or guest, speaks at a rate of more than 180 wpm (words per minute), it is very difficult to respeak all of the information. In fact, it has been established that the recommended speaking speed is 144 wpm. However, this recommendation is not usually fulfilled in debate programs and chat shows, due to their improvised nature. As the spoken word is one of their primary characteristics, these kinds of programs are not only full with

reformulations and false beginnings but also with grammatical or syntactic errors in some cases. In addition, there are other factors that make respeaking even more difficult including sudden and continuous changes of speakers, colloquial speaking styles background noises or bad pronunciation.

Because of all the inherent difficulties in the task, sometimes real-time subtitling does not achieve the quality of pre-recorded subtitling. In order to obtain better subtitling quality, the subtitles used in the experiment were not the live ones respoken for the first broadcasting but created specifically for the experiment. The subtitles were accordingly adapted to the six second rule (Díaz-Cintas: 2003), that states that for a maximum of six seconds on screen, subtitles should not have more than 72 characters (35 per line).

Table 6.1: Simplified table representing the Six Second Rule for subtitling

Seconds	Nº of characters	Seconds	Nº of characters
1	12	4	48
1'5	18	4'5	54
2	24	5	60
2'5	30	5'5	66
3	36	6	72
3'5	42		

Equally, the created subtitles tried to follow the Spanish standard rule UNE 153.010 "Subtitling for the deaf and hard of hearing: Subtitling through teletext."

Once the material was selected, the subtitles were created and inserted into the image. Afterwards, members from the Broadcast Technology Institute of Munich (IRT) slowed-down the clips into four different speeds: 90% playback speed (original speed slowed-down by 10%), 80% (original speed slowed down by 20%), 70% (original speed slowed-down by 30%) and 60% (original speed slowed-down by 40%). This final speed, however, was finally rejected for the experiment since not only the image but also the soundtrack were excessively affected by the manipulation. Thus, once the playback speed of the videos was manipulated, the length of the 32 clips was at follows:

Table 6.2: Duration of the clips used in the experiment

	CLIP 1	CLIP 2	CLIP 3	CLIP 4	CLIP 5	CLIP 6	CLIP 7	CLIP 8
100%	01:21	01:42	01:48	01:24	01:53	01:17	01:18	01:18
90%	01:29 **(+8")**	01:53 **(+11")**	02:00 **(+12")**	01.33 **(+9")**	02:06 **(+13")**	01:25 **(+8")**	01:27 **(+9")**	01:27 **(+9")**
80%	01:40 **(+19")**	02:07 **(+25")**	02:15 **(+27")**	01:44 **(+20")**	02:21 **(+28")**	01:35 **(+18")**	01:38 **(+20")**	01:38 **(+20")**
70%	01:54 **(+33")**	02:25 **(+43")**	02: 35 **(+47")**	02:00 **(+36")**	02:41 **(+48")**	01:49 **(+32")**	01:52 **(+34")**	01:52 **(+34")**

Since one of the evaluation criteria was comprehension, in spite of having manipulated the playback speed of the original video, the subtitles remained the same in the four samples. However, taking into account the already mentioned six seconds rule that governs the practice of subtitling, the slowed-down versions could have included an important number of extra characters in the subtitles:

Table 6.3: Extra number of characters after the playback speed reduction

		CLIP 1	CLIP 2	CLIP 3	CLIP 4	CLIP 5	CLIP 6	CLIP 7	CLIP 8
90%	Extra nº of seconds	+8"	+11"	+12"	+9"	+13"	+8"	+9"	+9"
	Extra characters	+96	+132	+144	+108	+156	+96	+108	+108
80%	Extra nº of seconds	+19"	+25"	+27"	+20"	+28"	+18"	+20"	+20"
	Extra nº of characters	+228	+300	+324	+240	+336	+216	+240	+240
70%	Extra nº of seconds	+33"	+43"	+47"	+36"	+48"	+32"	+34"	+34"
	Extra nº of characters	+396	+516	+564	+432	+576	+384	+408	+408

6.3.3 Experiment Design

Once the material was prepared and having selected the participants, the experiment commenced. As previously mentioned, participants watched the different videos with one of the eye trackers Tobii T60 belonging to TransMedia Catalonia research group. Each participant watched the eight different clips as well as the four different playback speeds. Therefore, each of them watched two videos in playback speed 100%, two in 90%, two in 80% and two in 70%.

Since comprehension, and not memory, was one of the main criteria to be evaluated through questionnaires, participants did not watch the eight clips one after the other, but in pairs of two. Equally, in order to control as much as possible factors such as short-term memory, tiredness or even lack of attention, the order of tasks of the experiment was meticulously prepared.

Thus, the experiment was divided into four different sections. Two participants took part in each section. Whereas one watched the videos in hearing conditions (with the previously mentioned three possible inputs activated), the other one watched the video in hearing under deaf conditions (auditory input deactivated). Each section was divided into four different tests, comprised of two videos. In order to avoid that these possible factors (short-term memory, tiredness or lack of attention) affect the results regarding the comprehension of each particular video, the position of the videos was carefully randomized. Thus, each pair of videos changed its position within each section (if Clip 1 and Clip 2 were within Section 1: Test 1, in Section 2 that pair of videos will not appear in Test 1 but in Test 2 and so on). Equally, the position of each video changed within each test (if in Section 1: Test 1 the order was Clip 1 and Clip2, in Section 2: Test 2 the order will be Clip 2 and Clip 1 and so on). Therefore, in such a randomized way, the already mentioned external factors will affect all the videos in the same way, without distorting the comprehension results. See Table 6.4 below for the distribution of the clips for the whole experiment.

Table 6.4: Distribution the video clips in the experiment

	SECTION 1		
Participants 1 and 2	**Test 1**	Clip: 1 Speed: 100%	Clip: 2 Speed: 70 %
	Test 2	Clip: 3 Speed: 80%	Clip: 4 Speed: 90%
	Test 3	Clip: 5 Speed: 80%	Clip: 6 Speed: 100%
	Test 4	Clip: 7 Speed: 70%	Clip: 8 Speed: 90%

	SECTION 2		
Participants 3 and 4	**Test 1**	Clip: 8 Speed: 80%	Clip: 7 Speed:100 %
	Test 2	Clip: 2 Speed: 100%	Clip: 1 Speed: 90%
	Test 3	Clip: 4 Speed: 80%	Clip: 3 Speed: 70%
	Test 4	Clip: 6 Speed: 90%	Clip: 5 Speed: 70%

	SECTION 3		
Participants 5 and 6	**Test 1**	Clip: 5 Speed: 100%	Clip: 6 Speed: 80 %
	Test 2	Clip: 7 Speed: 90%	Clip: 8 Speed: 70%
	Test 3	Clip: 1 Speed: 80%	Clip: 2 Speed: 90%
	Test 4	Clip: 3 Speed: 100%	Clip: 4 Speed: 70%

	SECTION 4		
Participants 7 and 8	**Test 1**	Clip: 4 Speed: 100%	Clip: 3 Speed: 90 %
	Test 2	Clip: 6 Speed: 70%	Clip: 5 Speed: 90%
	Test 3	Clip: 8 Speed: 100%	Clip: 7 Speed: 80%
	Test 4	Clip: 2 Speed: 80%	Clip: 1 Speed: 70%

6.3.4 Starting hypothesis

As it will be explained in the next section, one of the aims of the study was to triangulate observations in order to formulate a stronger hypothesis about the reception of slowed-down videos with subtitles. Thus, the starting hypotheses were related to both the data obtained through the comprehension questionnaires and the data obtained through the eye-tracking.

First of all, with regard to comprehension, it was expected that the slowed-down versions would have some effect on both the general understanding of the clip as well as on a more detailed comprehension. Whereas a general understanding was expected in the not manipulated videos, this was not only expected but somehow taken for granted in the slowed-down versions. On the other hand, and again in comparison to the not manipulated videos, a more detailed comprehension was expected in the slowed-down versions.

Secondly, regarding the eye-tracking analysis it was expected that the results of the slowed-down videos would show more fixations both in the image and in the subtitling area. Equally, as participants will have had more time to read the subtitles, it was expected that participants would go back more frequently when reading the subtitles, making more regressions.

The participants did not know that they were going to watch slowed-down videos. In this sense, it was expected that they would realize that, in the particular case of the most slowed-down version, that is the videos with a playback speed of 70%, the video had been manipulated. Since the effect of the manipulation was more evident in the audio than in the image, some different results between participants were expected. Thus, this assumption was almost taken for granted in the participants carrying out the experiment under hearing conditions and expected in the participants carrying out the experiment under deaf conditions. Equally, it was not expected that the participants would realize any manipulation in the videos with a playback speed of 80% and in the videos with a play back speed of 90%.

6.3.5 Analysis: triangulation

As it has been previously mentioned, in order to obtain stronger results about the reception of slowed-down videos with subtitles, the analytic evaluation was based on three different criteria: comprehension and eye-tracking analysis were the main criteria and opinion a secondary one. However, since the study was focused on a reception experiment, opinion had to be taken into account.

6.3.5.1 Eye –tracking

Eye-tracking, a technique used to record and measure eye movements, is aimed at identifying and analyzing patterns of visual attention when participants are performing a specific task. In this case they were simply asked to watch, as if they were at home, different subtitled clips from a TV program. The eye-tracker registers the eye movements of the participants and presents them to the

researcher in a meaningful and clear way. This is possible thanks to both the eye-tracker's near infrared illumination, which is reflected on the cornea and pupil of the user's eyes, and to its two image sensors, used to capture images of the participant's eyes and its reflection features. Equally, a first calibration measures the characteristics of the user's eyes and, together with the eye-tracker's 3D eye template, calculates the user's gaze data.

Two Tobii T60 eye-trackers were used in the experiment. Tobii T60 Eye Trackers are integrated into a 17" monitor. With an accuracy of 0.5 degrees, it transfers eye movement data to the computer every 16.6 milliseconds at a rate of 60 Hz.

Figure 6.2: Eye tracker Tobii T60

The eye-tracker collects two kinds of movement data: fixations and saccades. A fixation, represented by a dot, is the pause of the eye movement on a specific area of the visual field while a saccade, represented by a line, is the rapid movement between two fixations. Other movement data, relevant to this study, is the participant's gaze regression, that is, the movement made by the participant's eyes when s/he goes back in the reading of the subtitles and fixates their gaze in a previous point in the visual field.

Thus, having collected all this information, the eye-tracker can present the visual data either in a gaze plot or in a heat map. A gaze plot is the screen shot that shows all the fixations made by one participant on a specific video or image while a heat map is the screen shot that shows the total amount of fixations made by all the participants within certain areas of the video or image. The heat map uses colours to show the difference in the intensity of the fixations' length. Likewise, the eye-tracker allows the researcher to reproduce the video with all the fixations and saccades made by each participant.

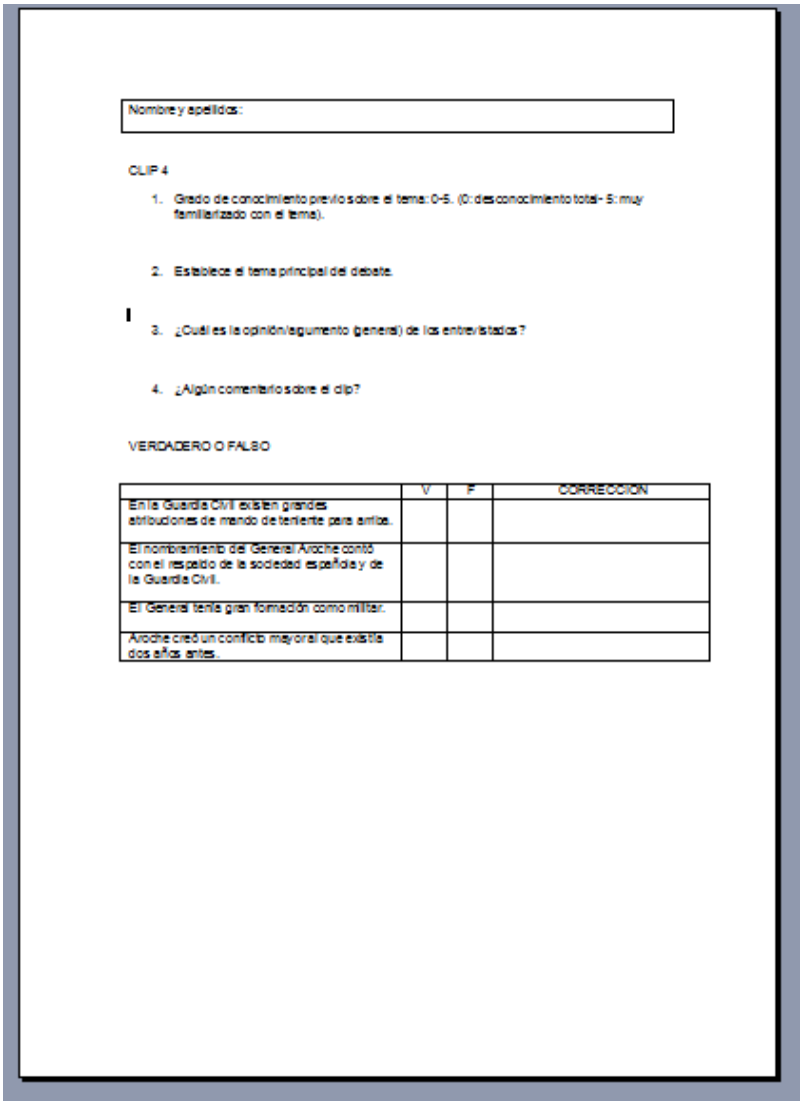
6.3.5.2 Comprehension questionnaire

As it has been advanced, one of the main objectives of the study was to assess if the comprehension of the videos improved in the slowed-down versions, especially when no auditory input was available. In order to prevent external factors interfering in the results, each test included a comprehension questionnaire. Thus, after watching two videos, the participant had to fill out a questionnaire.

Having a total amount of eight questionnaires, a questionnaire was prepared for every clip, regardless of the different playback speeds. All the questionnaires shared the same number of questions and overall structure. Each questionnaire was divided in two main parts: a first one aimed at assessing the general comprehension of the information presented in the clip, and a second one aimed at assessing detailed comprehension. See Figure 6.3 with a sample of the questionnaires.

The first part was made up of a closed question and three open questions. In the first question the participant was asked to choose within a scale their grade of knowledge about the topic discussed in the fragment of the video (being 0 “total ignorance” and 5 “highly familiarized with the topic”). In addition, they had the option to specifically state if they had already watched the program on TV, an important factor that would have a significant effect on the comprehension results. The following two questions were related to general comprehension. Thus, while in the first one participants were required to establish the topic discussed in the video, in the second one they were asked to specify the argument defended by each collaborator. No focus group was organized to establish valid answers because the videos were carefully chosen in order to present a unique topic for discussion (abortion, illegal downloading, bullfighting, Spanish Civil War and so on and so forth) as well as clear statements from the speakers in the debate, who took a clear position for or against the topic discussed. Finally, the participants were asked to make any pertinent remark about the quality of the video or any other relevant information. The data obtained in this question allowed the researcher to observe if the participants had realized or not that the videos had been manipulated.

The second part was made up of four true or false questions aimed at assessing detailed comprehension. Thus, in some questions specific data such as figures, names or dates were deliberately changed not only in order to observe if the participant remembered those pieces of detailed information but also to see if the participant was able to correct the false sentences.



Nombre y apellidos:

CLIP 4

1. Grado de conocimiento previo sobre el tema: 0-5. (0: desconocimiento total- 5: muy familiarizado con el tema).
2. Establece el tema principal del debate.
3. ¿Cuál es la opinión/argumento (general) de los entrevistados?
4. ¿Algún comentario sobre el clip?

VERDADERO O FALSO

	V	F	CORRECCIÓN
En la Guardia Civil existen grandes atribuciones de mando de teniente para arriba.			
El nombramiento del General Aroche contó con el respaldo de la sociedad española y de la Guardia Civil.			
El General tenía gran formación como militar.			
Aroche creó un conflicto mayor al que existía dos años antes.			

Figure 6.3: Questionnaire

6.3.5.3 Opinion Survey

Once the participants had watched all the eight videos and answered the corresponding questionnaires, the aim of the experiment was explained to them. Therefore, they could express their opinion regarding the different playback speeds and their effect on the audio, image as well as on the comprehension of the clip.

Since groups in risk of exclusion do not want to suffer neither negative nor positive discrimination, the opinion of the participants was taken into consideration when drawing the final conclusion, in order to choose the percentage of speed reduction that best fit potential users' needs.

6.4 Obtained Results

In this section the results of the whole experiment will be offered, followed by a brief discussion of them. Within the framework of DTV4All Emerging Access Services, the main objective of the discussion is to introduce new research possibilities regarding reducing playback speed as a tool for ensuring accessibility to TV contents.

6.4.1 Eye-tracking results

The eye-tracking data was obtained from three different tools for visualizing the gaze information: statistical charts, gaze plots and heat map plots.

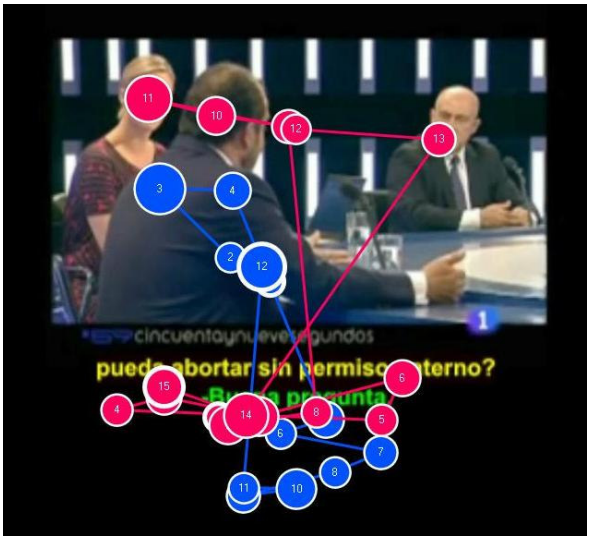
The statistical chart is based on the length of participants' fixations into the two established Areas of Interest (AOI): AOI_1 (subtitling area) and AOI_2 (rest of the image), see Table 6.5.

Table 6.5: Statistical chart of Video 1-Playback speed 100%: Area of interests and length of fixations

1-100	AOI_1 (seconds)	AOI_2 (seconds)
Ander	1.732193	1.764037
Aitor	2.835151	1.548687
Min	1.732193	1.548687
Max	2.835151	1.764037
Sum	4.567344	3.312724
Mean	2.283672	1.656362
Median	2.283672	1.656362

The chart offers the following information: min. (minimum fixation length), max (maximum fixation length), sum (sum of all fixation lengths), arithmetic mean (sum of all fixation lengths divided by the number of values in the data set) and median (mid point when ordering the fixation lengths). Since one of the main objectives was to compare the time participants fixated on the image and the time they fixated on the subtitles area, no more specific areas of interest were delimited.

As far as the gaze plot is concerned, the gaze plot is a statistic way of visualizing the gaze data of each previously selected sequence of images in the same shot, see Figure 6.4.



Picture 6.4: Gaze plot of Video 1-playback speed 100

In the gaze plot, each fixation is illustrated with a coloured dot. On the one hand, the radius represents the length of the fixations and, on the other hand, the number which appears inside the dot represents the order in which fixations were made.

Finally, a heat map plot presents the stimuli (the first image of the selected sequence of images in the same shot) as the background image and a heat map mask is superimposed on top of it. The highlighted areas are the areas which participants have been looking at. Three different kinds of heat map were studied in the experiment: a heat map with the gaze data of the two participants carrying out the same section, another heat map with data exclusively corresponding to the participant that carried out the experiment under hearing conditions and a final heat map with the gaze data of the participant that watched the videos under deaf conditions, see Figure 6.5.

Figure 6.5: Gaze maps of Video 1, playback speed 100%. From left to right: all participants-under hearing conditions-under deaf conditions

Before summarizing the results obtained from these visual sources of information, it has to be mentioned that in some images a kind of visual imbalance could be appreciated between the subtitling area and the participants' fixations due to sudden changes in both participants' body and head position. Those position changes could affect the way they looked at the screen. However, having selected simple and big areas of interests, this fact does not have any repercussions when it comes to analyzing the data, since the two areas are very easily identified.

Interesting gaze behaviors were observed in the eye-tracking analysis of the eight participants. These behaviours are related to ascendant progression expected in the fixation length, to gaze dispersion and new focus of interest in the slowest versions and to the differences found between groups of participants.

An ascendant progression (100% playback speed <90% playback speed <80% playback speed <70% playback speed) in the length of fixations in both areas of interests was obviously expected. In other words, it was expected that fixation length increased in a regular and progressive way from the not manipulated speed to the most reduced one. However, not such a clear pattern has been found within all the videos. In fact, only Video 1 shows a clear ascendant pattern in both participants fixation length in both areas of interests. In the rest of the clips, differences have been found in the fixation length of each area of interest. Thus, and regarding to AOI1 (subtitling area), although a clear ascendant pattern between the 100% playback speed and the 70% playback speed has been found, within the rest of speeds the distribution turned out to be quite irregular. For instance, analysis of Video 2 showed that 80% playback speed gathered less fixation length in AOI1 than the 100% and 90% playback speed versions. However, with regard to AOI2 (rest of the image) a clear ascendant progression has been found in almost all the videos. A possible response to this irregular pattern could be found in the following interesting gaze behaviour detected: the difference between the fixation lengths of both areas of interests is much more balanced in most reduced playback speeds. In other words: in 100% and 90% playback speed the difference between the time participants fixated on the subtitling area and the image was more accentuated that in the 80% and 70% versions. In some 70% speed cases, the time spent in both areas was almost similar (See the statistical charts of Video 5, Video 6 and Video 7) whereas in 100% playback speed the highest percentage of fixation length was on the subtitling area. Equally, as can be clearly appreciated in the charts for Video 1 and Video 4, fixation length in the subtitling area tends to stop increasing at the 80% playback speed. In this sense, two clear conclusions could be drawn: on the one hand, in 100% playback versions, the participants focused the greater part of their attention on the subtitling area since they had no time to fixate in both image and subtitles. On the other hand, it could be equally

concluded that once the subtitles have been read, not much time is spent in fixating in this area but in the image instead.

With regard to the gaze dispersion, this phenomenon was surprisingly found in some of the 100% and 70% playback speed versions and repeated throughout the eight videos. Whereas gaze dispersion in 100% playback speed versions can be attributed to the excessively quick subtitles, gaze dispersion in 70% playback speed versions could be attributed to the excessively slow subtitles. In the former, fixations are mainly concentrated on the subtitling area maybe due to the fact that participants did not have much time to focus on both areas of interests, so they moved their gaze across the image in quite a chaotic way. (See eye-tracking results of Video 6). In the latter, the great amount of both regressions in the subtitling area and continuous gaze movements from subtitles to image (and vice versa) could mean that participants had too much time to read the subtitles so they continuously moved across the image. (See eye-tracking results of Video 7)

As far as the appearance of new focus of interests is concerned, this phenomenon could be easily found in some of the 70% playback speed versions. For instance, in Video 3 fixations on a bull appearing on a screen behind the presenter could only be found in the reduced versions. Equally, in Video 5 fixations on the face of both participants are only found in the reduced slowest versions, whereas in the not manipulated speed participants only fixated on the face of the first speaker.

Finally, not many differences were found between participants carrying out the task under hearing or deaf conditions. Reducing the playback speed of the videos has turned out to help all viewers equally, regardless of the presence or absence of the auditory input, and in both comprehension and visual processing. In this sense, the sample size was reduced (four participants with auditory input and four without it), so further research could be recommended in order to give more light into the perception differences between not only hearing and deaf participants, but also between cognitive impaired people, old people and the rest of potential groups that could benefit from slowed down versions.

6.4.2 Comprehension results

Regarding comprehension, the following chart will summarize the results obtained in the eight different questionnaires filled out by the participants.

Except in one case, marked with an asterisk, none of the participants had previously watched the TV programs from which the videos were extracted. Thus, the chart gathers the participants' comments on the clips, exclusively the ones which were related to either bad synchrony or an

excessive slowdown possibly caused by the manipulation they suffered, together with the obtained results in both general and detailed comprehension. Concerning general comprehension, one of two possible marks could be given: a right answer (marked with a ✓) or a wrong answer (marked with an X). Concerning detailed comprehension, and taking into account that it was a true or false exercise, a third possible mark was added: "OK without correction", referring to the false sentences that were identified as false but were not corrected by the participant with the right answer.

Table 6.6

VIDEO 1 Hearing conditions	100%	90% speed	80% speed	70% speed
Comments	No	No	No	No
General comprehension	✓	✓	✓	✓
	✓	✓	✓	✓
Detailed comprehension	OK without correction	✓	✓	✓
	✓	X	✓	✓
	✓	✓	✓	✓
	X	✓	✓	✓

VIDEO 1 Deaf conditions	100%	90% speed	80% speed	70% speed
Comments	No	No	No	Slowed-down image
General comprehension	✓	✓	✓	✓
	✓	✓	✓	✓
Detailed comprehension	X	✓	Ok without correction	✓
	X	X	X	✓
	Ok without correction	✓	Ok without correction	✓
	✓	✓	✓	✓

VIDEO 2 Hearing conditions	100%	90% speed	80% speed	70% speed
Comments	No	No	No	Slowed-down image
General comprehension	✔	✔	✔	✔
	✔	✔	✔	✔
Detailed comprehension	✔	✔	Ok without correction	✔
	✔	✔	✔	✔
	X	✔	Ok without correction	✔
	✔	✔	✔	✔

VIDEO 2 Deaf conditions	100%	90% speed	80% speed	70% speed
Comments	No	No	No	No
General comprehension	✔	✔	✔	✔
	✔	✔	✔	✔
Detailed comprehension	✔	Ok without correction	✔	Ok without correction
	✔	✔	✔	✔
	✔	X	✔	X
	X	Ok without correction	X	Ok without correction

VIDEO 3 Hearing conditions	100%	90% speed	80% speed	70% speed
Comments	No	No	Bad synchrony	Bad synchrony
General comprehension	✔	✔	✔	✔
	✔	✔	✔	✔
Detailed comprehension	✔	✔	✔	✔
	✔	Ok without correction	Ok without correction	✔
	✔	✔	✔	✔
	✔	✔	✔	✔

VIDEO 3 Deaf conditions	100%	90% speed	80% speed	70% speed
Comments	No	No	No	No
General comprehension	✔	✔	✔	✔
	✔	✔	✔	✔
Detailed comprehension	✔	✔	✔	✔
	Ok without correction	✔	Ok without correction	✔
	✔	✔	✔	✔
	Ok without correction	✔	Ok without correction	✔

VIDEO 4 Hearing conditions	100% speed	90% speed	80% speed	70% speed
Comments	No	No	No	Slowed-down image and bad synchrony
General comprehension	✓	✓	✓	x
	✓	✓	✓	X
Detailed comprehension	X	X	X	X
	✓	X	X	X
	X	✓	✓	✓
	X	✓	✓	✓

VIDEO 4 Deaf conditions	100%	90% speed	80% speed	70% speed
Comments	No	No	No	No
General comprehension	✓	✓	✓	✓
	✓	✓	✓	✓
Detailed comprehension	X	Ok without correction	X	X
	✓	X	X	✓
	✓	✓	✓	✓
	X	✓	✓	X

VIDEO 5 Hearing conditions	100%	90% speed	80% speed	70% speed
Comments	No	No	No	No
General comprehension	✓	✓	✓	✓
	✓	✓	✓	✓
Detailed comprehension	X	✓	✓	X
	X	✓	Ok without correction	Ok without correction
	✓	✓	✓	✓
	✓	X	✓	✓

VIDEO 5 Deaf conditions	100%	90% speed	80% speed	70% speed
Comments	No	No	No	Slowed-down image
General comprehension	✓	✓	✓	✓
	✓	✓	✓	✓
Detailed comprehension	X	✓	X	✓
	Ok without correction	✓	Ok without correction	✓
	✓	✓	✓	✓
	X	✓	X	✓

VIDEO 6 Hearing conditions	100%	90% speed	80% speed	70% speed
Comments	Quick subtitles	No	Synchrony	Synchrony
General comprehension	✓	✓	✓	✓
	✓	✓	✓	✓
Detailed comprehension	Ok without correction	✓	X	X
	X	✓	X	X
	✓	✓	✓	✓
	✓	✓	✓	✓

VIDEO 6 Deaf conditions	100%	90% speed	80% speed	70% speed
Comments	No	No	No	Slowed-down image
General comprehension	✓	✓	✓	✓
	✓	✓	✓	✓
Detailed comprehension	X	✓	Ok without correction	X
	X	✓	✓	✓
	✓	✓	✓	✓
	✓	✓	✓	X

VIDEO 7 Hearing conditions	100%	90% speed	80% speed	70% speed
Comments	No	No	No	No
General comprehension	✓	✓	✓	✓
	✓	✓	✓	✓
Detailed comprehension	X	✓	✓	✓
	✓	✓	X	✓
	✓	✓	✓	✓
	Ok without correction	✓	✓	✓

VIDEO 7 Deaf conditions	100%	90% speed	80% speed	70% speed
Comments	No	No	No	No
General comprehension	✓	✓	✓	✓
	✓	✓	✓	✓
Detailed comprehension	X	Ok without correction	✓	Ok without correction
	✓	✓	✓	✓
	✓	✓	✓	✓
	X	Ok without correction	✓	Ok without correction

VIDEO 8 Hearing conditions	100%	90% speed	80% speed	70% speed
Comments	No	No	No	Slowed-down image
General comprehension	✓	✓	✓	✓
	✓	✓	✓	✓
Detailed comprehension	✓	✓	Ok without correction	✓
	✓	✓	✓	✓
	X	✓	X	X
	✓	✓	✓	✓

VIDEO 8 Deaf conditions	100%	90% speed	80% speed	70% speed
Comments	Quick subtitles	No	No	No
General comprehension	✓	✓	✓	✓
	✓	✓	✓	✓
Detailed comprehension	✓	Ok without correction	X	Ok without correction
	✓	✓	✓	Ok without correction
	X	X	✓	Ok without correction
	✓	✓	✓	✓

In terms of general comprehension, as it can be appreciated in the chart, there is a clear uniformity within all the videos and all the playback speeds. In this sense, all the participants were able to identify the topic of discussion as well as to identify the stance taken by both debate participants. The only exception was found in a participant when watching a 70% playback speed video under hearing conditions. As he pointed out, he soon realized that the video had been manipulated as it was excessively slow. Consequently, he got distracted and directed his attention to the quality of the video and not to the content that was being transmitted. That is the reason why he could establish neither the topic of the discussion nor the opinion of the participants.

The results regarding the detailed comprehension, on the contrary, are not so uniform, although a clear conclusion could be drawn: as expected, all the three reduced playback speeds (90%, 80% and 70%) obtained better detailed comprehension results than the not manipulated playback speed (100%). However, in relation to the participants' comprehension behaviour within each particular video, no clear progression line could be drawn. Contrary to expectations, the highest playback speed reduction (70%) did not always give the best comprehension results. In this sense, there were several external factors that made the task of assessing detailed comprehension even more complicated. For instance, although all the participants were asked not only to choose between true or false but also to correct the wrong sentences, two participants did not make any correction throughout the eight questionnaires. However, despite these factors (or maybe because of them), the conclusion remains equally clear: greater playback speed reduction in the videos does not always result in better comprehension results obtained by the participants.

In fact, better results were obtained in the 90% playback speed than in the 70% playback speed. Each participant watched two videos of each speed category. Therefore, a total of 16 questionnaires were focused on assessing the comprehension of each percentage slow-down. Taking into account only the number of times in which all the questions were correctly answered (and corrected), the results were as follows:

- 90%: 8 questionnaires were correctly answered (and corrected).
- 70%: 7 questionnaires were correctly answered (and corrected).
- 80%: 2 questionnaires were correctly answered (and corrected).
- 100%: 1 questionnaire were correctly answered (and corrected).

Note: 'Corrected' in this sense refers to the false statements that were identified as false and were corrected by the participant with the right answer.

As has been mentioned, surprisingly there was not a standard progression in comprehension. A possible response to this irregular pattern was looked for in the participants' comments on the clips.

Thus, regarding the 70% playback speed, in 6 questionnaires out of 16 the participants made some remark about the bad synchrony or the excessive slowdown of the video. Once they had finished the experiment, the participants were asked about these particular remarks and their response was quite uniform and categorical: the excessive slowdown both bothered and distracted them. In this sense, the reason why the 70% playback speed did not obtain the best results could be found in this fact: if the viewer perceives the manipulation, excessive reduction of the original playback speed can be counterproductive. However, this response would not clarify the reason why the 80% playback speed obtained such bad results, since only two participants remarked on an excessive slowed-down speed within this category. Regarding the 90% playback speed, where the best results were obtained, none of the participants noticed any kind of manipulation.

Thus, a possible response to this participants' behaviour could be found in the subtitles. Since no extra characters were added in the reduced playback speed (the subtitles remained exactly the same in all the reduction percentages) it is probable that participants had excessive time to read the subtitles and, consequently, got distracted more easily. Thus, although participants did not perceive any manipulation in the 80% playback speed, the subtitles maybe still stayed too long on screen without adding new information, so participants stopped paying attention and obtained counterproductive results. In this sense, the playback speed with better results, not only in terms of comprehension but also in terms of less intrusive manipulation, is the 90% playback speed.

Taking into account these obtained results, however, it would be interesting to continue doing further research. For instance, as was established in Tables 6.2 and 6.3 reducing the playback speed could offer the opportunity to increase the amount of information being subtitled by adding more characters. Thus, further research could be done in order to establish a balance between the reduction of playback speed and the addition of new information. Given that both viewers under both hearing and deaf conditions, perceived the manipulation in the 70% playback speed, it would be interesting to focus exclusively on 80% playback speed and compare it with the 90% playback speed. In order to observe if better results could be obtained if more information is added to the subtitles another experiment could be carried out. Following the data gathered in Tables 6.2 and 6.3, it could be concluded that: given that 1'30 seconds is the average duration of the videos, a slowed-down speed of 80% could allow the addition of an average of 22 extra characters. This estimation, based on the 6 second rule would correspond to a standard subtitling reading speed. Thus, in order to facilitate access to those with difficulties in standard reading, hearing or understanding, the number of characters could be reduced to 10-12 extra characters. In order to establish the optimum slow-down speed percentage, the results could be compared with the playback speed of 90% in which no extra characters would be added.

On the other hand, the only significant difference found between the participants that accomplished the task under hearing conditions and the ones that carried out the experiment under deaf conditions was that the latter did not perceive the manipulation as clearly. In this sense, it could be concluded that slow-down reduction has more effect on the audio than on the image. Consequently, by not receiving the auditory input, participants under deaf conditions were less aware of the manipulation. In fact, within the questionnaires corresponding to the 32 clips shown without audio, only two remarks about speed reduction were found. Both in the 70% playback speed.

Finally, another interesting issue was to compare the number of wrong answers and the number of not remembered corrections. In the not manipulated videos, a total of 22 questions were wrongly answered. In contrast, in the 90% playback speed, the number of wrong answers dropped considerably to eight. In the 80% playback speed, 12 questions were wrong and 11 in the case of 70% playback speed. Therefore, since the number of corrected answers clearly improved, it could be again established that slowing down the playback speed has a clear effect on the comprehension of the clip. However, with relation to the "OK but without correction" answers, the obtained results in all the reduction percentages followed a similar pattern: seven questions were left without the right correction in the 100% and 90% playback speed; eight, in the 70% playback speed and eleven, in the case of 80% playback speed. Although here again the 90% playback speed obtained the best results, no considerable differences were found between reduction percentages.

In short, further research is necessary in order to establish better criteria for reducing playback-speed for accessibility, but several conclusions could be drawn: playback speed reduction clearly has a positive effect in viewers' comprehension. In addition, and surprisingly, it could be concluded that the percentage with better comprehension results is the 90% playback speed and not the 70% playback speed, as expected. Thus, a not too large speed reduction is needed in order to improve comprehension.

6.4.3 Opinion results

After the experiment, the aim of the study was explained to the participants. They were told that the videos had been displayed in four different speeds in order to check if the videos with most reduced speed were better understood. Thus, they could express their opinion regarding the different playback speeds and its effects on the audio and the image as well as on the comprehension of the clip.

Almost all of the participants, both under hearing and under deaf conditions, made some remark about the quality of the video. In this sense, the selected clips were extracted from TV programs so

the quality of the image was not high. However, regarding manipulation detection, there was a clear difference between participants that carried out the task under hearing conditions and the ones that watched the videos under deaf conditions. Thus, the great majority of the former (three out of four) made specific remarks about speed reduction. One of them clearly identified that some videos were slowed down (the videos with 70% playback speed), and the other two identified some manipulation in the general sound and in the voices of both TV host and debate participants. Once the aim of the study had been explained, all of them assured us that they did notice some slow down in the videos, but only in the most reduced ones. One of them concluded that, in opposition to the not manipulated videos, the rhythm of the subtitles was more easily followed in the slowed-down videos. However, all of them agreed that, once some manipulation was identified excessive speed reduction was counterproductive. On the other hand, among the four participants under deaf conditions only one made a specific remark about playback speed reduction (in the case of playback speed 70%). He categorically declared that the image was very slow and, consequently, it was "quite exasperating" reading the subtitles. Once explained the aim of the study, only two of them assured us that they did clearly notice some slow down in the videos. The other two participants did not notice any playback speed reduction but declared that they did realize that it was "surprisingly" very easy to read the subtitles and simultaneously watch the video.

Two clear conclusions can be drawn from the participants' opinions: on the one hand, they did declare that, in comparison to the not manipulated videos in which subtitles disappeared quite quickly, the reduced videos were more easily followed. However, when the speed reduction is evident in the image, and specifically, in the audio quality, it became "uncomfortable" to watch the clips.

6.5 Conclusions

Conclusions about the effect of reducing playback speed as a tool for accessibility, widely disseminated throughout the previous sections, are compiled and summarized in this section. Thus, although the conclusions have already been presented within each of the sections corresponding to the three established analysis criteria (visual processing through eye-tracking, general and detailed comprehension through questionnaires and opinion through opinion surveys), they are here presented from a consolidated point of view.

First of all, results have clearly shown that reducing playback speed does have an improvement effect in both visual processing and comprehension.

In terms of visual processing, eye-tracking analysis led us to conclude that playback speed reduction helps viewers to have more time to not only read the subtitles but also to pay more attention to the image. However, excessive playback reduction could be counterproductive: whereas in 100% playback speed versions participants' gaze could be disperse due to excessive quick subtitles, in 70% playback speed versions the same phenomenon could also take place if the subtitles (and both image and audio) are excessively slow.

Although in terms of general comprehension, there is a clear uniformity within all the speed percentages, in terms of detailed comprehension, the same conclusions can be drawn. On the one hand, as expected, all the reduced playback speeds obtained better results than the not manipulated playback speed. However, on the other hand, a greater reduction in playback speed does not always mean better comprehension results. In other words: if the viewer perceives the manipulation, excessive reduction of the original playback speed can be counterproductive.

Finally, and taking into account participants' opinions, another two clear conclusions can be drawn: whereas quick subtitles do make it more difficult to both watch and understand a particular video, excessively slow images and subtitles do make it more difficult to focus on the video and its subtitles since evident speed reduction distracts participants.

In short, once proved that playback speed reduction has a positive effect in both viewers' comprehension and visual perception, further research could be done in order to shed more light on different factors, preliminarily treated in the current experiment.

Thus, for instance, more reception experiments could be carried out in order to study the possible differences (if any), in both visual perception and comprehension, between the different groups of viewers that could potentially benefit from slowed-down video: deaf and hearing impaired viewers, older people, cognitively impaired people and so on and so forth. Equally, it would be interesting to continue doing further research in order to establish better criteria for reducing playback-speed for accessibility, such as establishing a balance between the reduction of playback speed and the addition of new information by adding extra characters to the slowed-down subtitles. Finally, another interesting idea would be to apply slowed-down playback speed in TV or DVD programs as a useful tool in second and third language acquisition.

7. Tests with a Demonstrator of Enhanced Audio Description

By Jordi Mata / Jordi Arraez / Xavier Vives Surroca

Partners involved: TVC, UAB

7.1 Introduction

Audio Description (AD) is an additional audio track with narration for blind and visually impaired people. This service can be delivered using different techniques, over DVB-T with the help of a second pair of audio channels (broadcast/broadcast) or with the help of a shared distribution, broadcast and broadband via the air and the web (called a hybrid solution), even a fully web based solution can be considered called WebTV. Because the AD audio track can be delivered using technologies that are not currently tested for this purpose (see synchronisation issues in broadcast/broadband distribution) it is regarded as an emerging access service.

7.2 Description of the implemented technique

TV Catalonia has established an almost regular service containing AD with the help of DVB-T transmitters in the local area around Barcelona. People suffering vision impairments can switch to an additional narrative sound channel explaining what is going on. TVC has established a feedback service with users both at association level and with individuals who write regularly offering their opinions of the service. Tests with users have been undertaken by TVC and UAB as part of the work of the project on "mature services" and the results will be available in D2.4.

It is expected that the results of the user interviews which have been undertaken by TVC and UAB, and are reported in D2.4, concerning AD are true and valid for WebTV and Hybrid-TV AD. In other words, the experience of the viewers is supposed to be the same in each case as the presentation of AD is completely independent of the way it is delivered. However, several tests will be conducted in order to validate this assumption. Section 7.2 outlines the kind of tests that will be conducted by UAB and TVC, although the results will be later described in deliverable D3.6.

In the current section, we will explain the 5 scenarios of emerging services that deal with audio description:

1. Live streaming Internet TV: TV contents with an AD channel (broadcast mix) are streamed via IP.
2. On demand AD: Three scenarios centred on the consumption of recorded contents. These scenarios let users' access contents at any time, not just at the time when they are

broadcasted by Digital Terrestrial Television (DTT). Three different scenarios deal with on-demand AD:

a. Video On-Demand with AD using a Set-Top-Box: Video On-Demand by means of IP streaming.

b. Content downloading with AD for consumption on multimedia devices: Video On-Demand by means of downloading video and audio files.

c. Audio description files downloading (Podcast): Video On-Demand by means of downloading only the AD audio files.

3. Simultaneous AD and non AD consumption: No IP distribution here. Instead, DTT contents are simultaneously consumed in a place (normally a family situation) where visually impaired and non-impaired people consume the same contents at the same time in the same place.

These are emerging scenarios and, therefore, are not currently being exploited. TV Catalonia expects to deliver 5 prototypes, one per scenario that will be used to demonstrate the feasibility or otherwise of each of these 5 scenarios and that will be used for testing purposes.

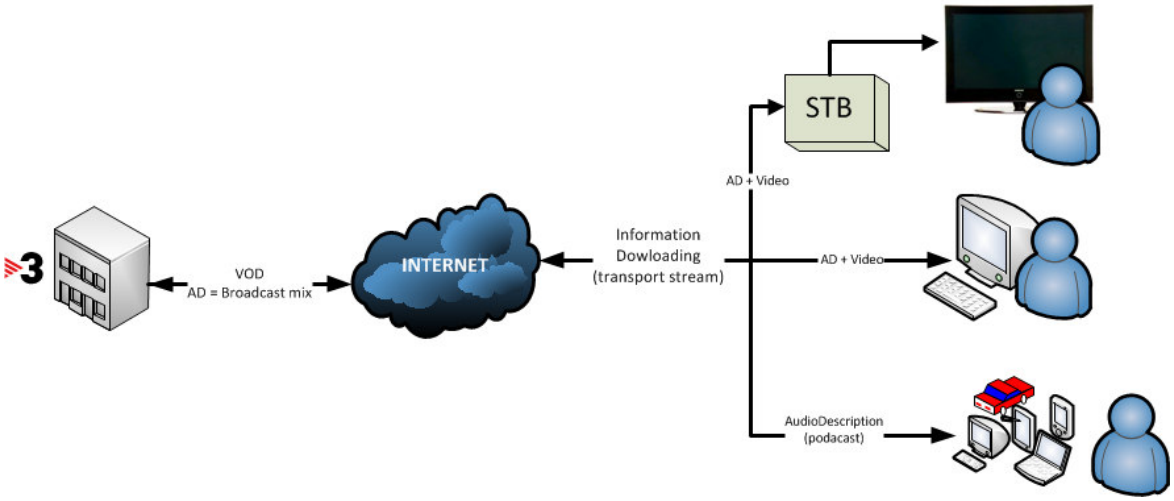


Figure 7.1: Schema of the 3 On-Demand scenarios: delivery of recorded contents

7.2.1 Live streaming Internet TV Emerging Scenario

The *Live Streaming Internet TV scenario* is a prototype that emulates DTT broadcasting, but by means of the Internet uses an IP channel instead of free to air broadcasting. The following diagram represents the distribution channel:

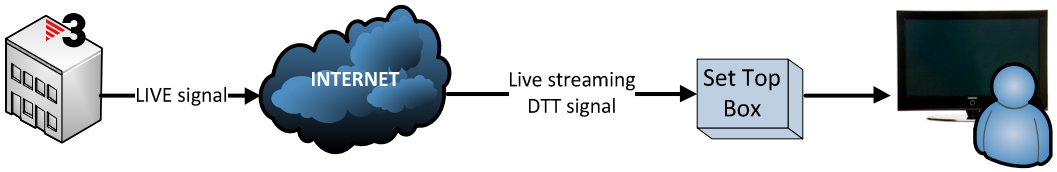


Figure 7.2: *Live Streaming Internet TV* distribution

The emitted live signal will include a high quality video signal, and a minimum of two audio signals, which are:

- Stereo Catalan audio channel
- Mono Audio Description channel: it will be a mix of the normal audio and the audio description. This mix will be performed at broadcaster's side.

For the correct reception and decoding of the live streaming, the end user will need to use a properly configured set top box. TVC will use a hybrid DTT/IP set top box named *Netbox 8160*, which is manufactured by the company *Netgem.*

Figure: 7.3: Set top box used for the prototypes (Netbox 8160)

7.2.1.1 Preparation of the IP contents for live broadcasting

Technical features

Live Streaming emissions will follow the DVB-IP standard. This standard deals with the encapsulation in TC/IP packages of the *Transport Stream* packets that are used in DTT. The content encoding is performed with the following characteristics:

- Video component: TVC will use a H.264 codification with an output coding rate of 1.5Mbps. This kind of codification allows end users to view an image with a quality that is similar to the one that is broadcast in free to air DTT.
- Catalan audio component: It is codified with Advanced Video Coding (AAC) with a bit rate of 128 Kbps.

- Audio description component: It is also codified with AAC with a bit rate of 128 Kbps.

Synthesis of the live feed contents for IP

The source signal that is used in the synthesis of IP content for live broadcasting content is the same signal that feeds the DTT header. It is originally delivered over a coaxial BNC (Bayonet Neill-Concelman) cable with an audio-embedded Serial Digital Interface (SDI) format. A PC with a *BlackMagic DeckLink SDI PCIe* model board inside will be used for the prototype. The coaxial cable with the SDI input signal will be connected to the *BlackMagic* board. The PC will be used to run software that will codify the original signal into a DVP-IP stream that will be sent as a unicast signal to the Real Time Streaming Protocol (RTSP) publication module. Figure 7.4 represents the content synthesis for the IP live broadcasting process:

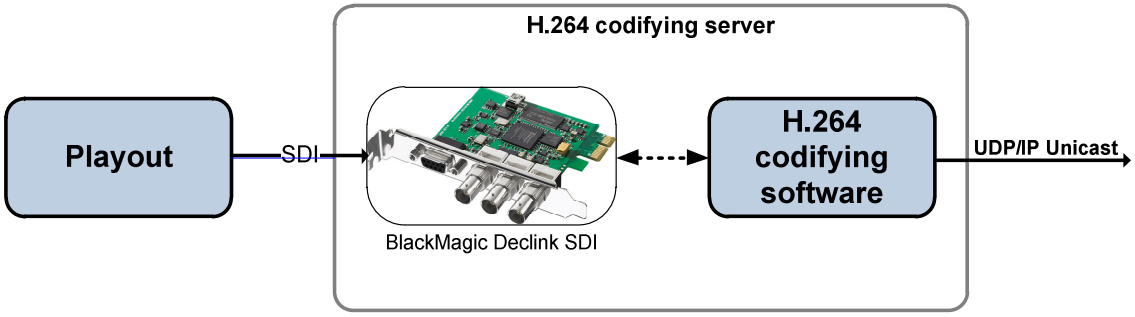


Figure 7.4: Synthesis of the live feed contents for IP

7.2.1.2 System for IP live broadcast contents publishing

RTSP Server for live broadcasting

The RTSP server for live broadcasting gathers on one side the unicast streaming with the live contents served by the H.264 encoding server, and outputs on the other side the content that users request.

The protocol that users' side set top boxes follow for demanding live signal is the RTSP protocol through port 554. This protocol lets the set top box establish, maintain and close a User Datagram Protocol (UDP) channel. Thanks to this UDP channel, the server is able to provide the live streaming that will be provided to each user individually.

Figure 7.5 shows this part of the process:

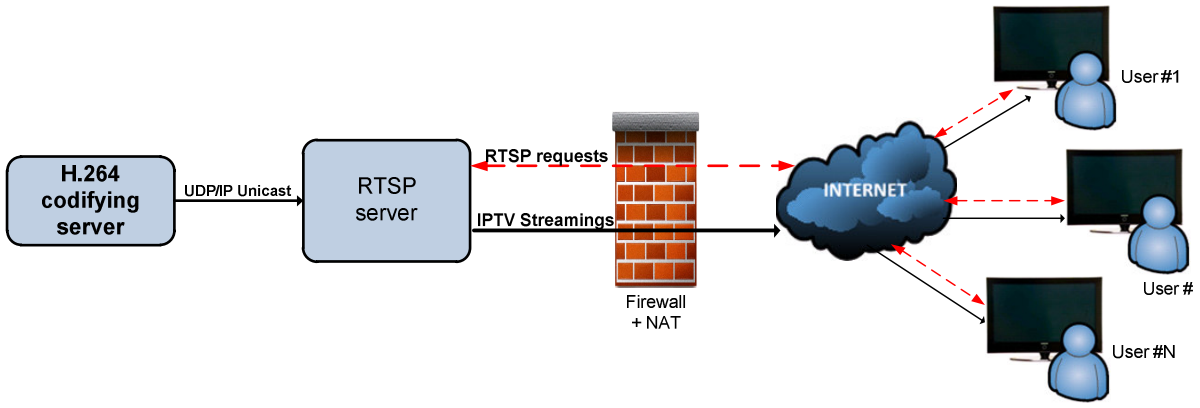


Figure 7.5: RTSP server for live IP broadcasting

Opening the system to Internet

The users of the prototype access the RTSP server via Internet thanks to the gateways that have been activated by TVC's Information Technology (IT) Department personnel. The gateways will enable a way through the firewall that controls the access to the Internet throughout the corporate intranet.

To do so, on one side the requests of the RTSP protocol (port 554) of the Internet test users to a corporate IP address have been enabled. These requests are forwarded to the RTSP server thanks to a NAT (Network Address Translation) technique. On the other hand, the IT Department has reserved and kept enough bandwidth on the connection output to ensure a proper viewing of live streaming broadcasts on the client side.

7.2.1.3 Users' setup

Receiver software

The NetBOX hybrid receiver that is used in the prototype is able to give access to a website that lets users choose between the two available audios channels: Catalan and audio description. Once the user selects the desired audio, the receiver automatically tunes the live IP channel and unicasts the audio that has been selected by the user.

The following screenshot shows the audio selection menu that lets users choose the audio channel:

Figure 7.6: Audio selection menu (IP live streaming)

7.2.2 Video On-Demand AD content over a Set-Top-Box

7.2.2.1 Description of the scenario

The Video On-Demand AD content over a Set-Top-Box scenario lets the user select contents on demand from a list of programs. S/he will then be able to watch these contents on the TV screen. The contents will be broadcast via the Internet thanks to a streaming IP channel. The following figure shows the distribution channel:

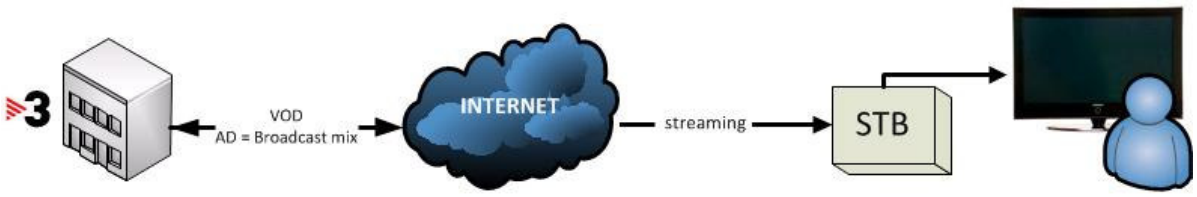


Figure 7.7: VoD with AD (broadcaster mix) distribution

The clips that will be available for the test will contain a high quality video and two audio channels:

- One stereo Catalan channel
- A Mono audio description channel that is a mix of the normal audio channel plus the mixed audio description (the mix is done at the broadcaster side).

For the IP streaming Video On-Demand contents, the user will use the following set top box. It is a hybrid model DTT/IP: Netbox 8160 of *Netgem* manufacturer. It is the same as that shown in Figure: 7.3.

7.2.2.2 Preparing the contents for IP Video On-Demand

The archive of programs for Video On-Demand (the so-called Catch-Up) is a set of files that are generated by means of transcoding the regular DTT broadcasts of TV3 to *Transport Stream* packets. This format will later enable a distribution that is in streaming mode and follows the DVB-IP standard.

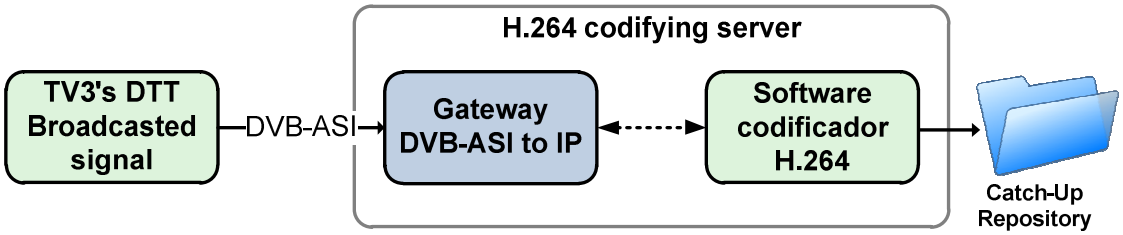


Figure 7.8: AD enriched VoD generation

The content codifying is done the following way:

- Video component: content is codified with H.264 with an output codifying rate of 1.5Mbps. This codification provides a high quality image that is similar to the one of a DTT broadcasted signal.
- Catalan audio component: It is codified with AAC with a bit rate of 128 Kbps.
- Audio description component: It is also codified with AAC with a bit rate of 128 Kbps.

All the files that have been created for Video On-Demand will be saved in a repository and will be available to be served.

7.2.2.3 Publishing system: RTSP server for VoD

The RTSP server for Video On-Demand supplies via *streaming* the contents that are available in the Catch-Up repository whenever users ask for them.

The protocol that is followed by users' receivers whenever they are demanding content is the RTSP protocol via port 554. This protocol lets equipment establish, maintain and close the UDP channel that will is used to individually supply the streaming to each of the end-user receivers.

Figure 7.9 represents this part of the process:

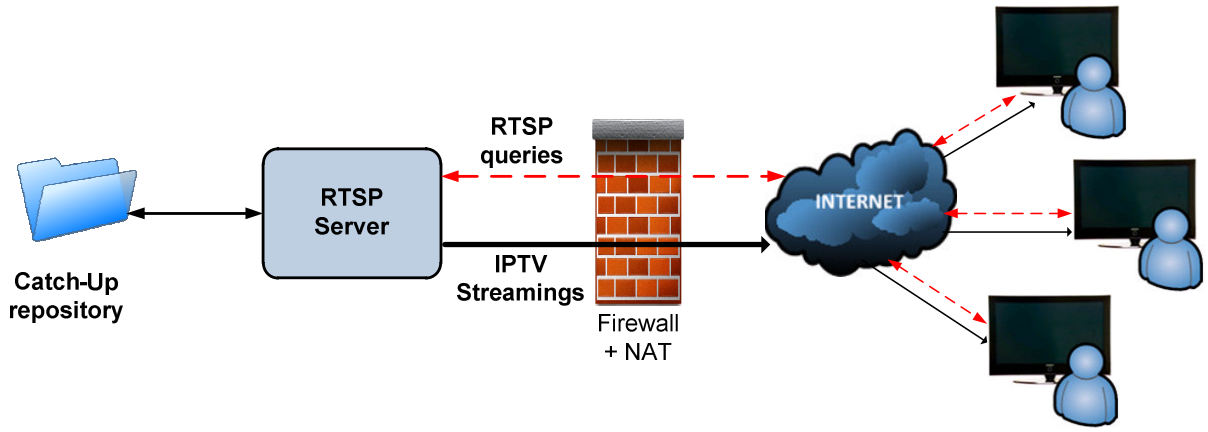


Figure 7.9: RTSP Server for VoD broadcasting

7.2.2.4 Opening the system to Internet

The users of this emerging AD service access the RTSP server from the Internet, thanks to the gateways that TVC's IT Department has enabled in the firewall that controls the access to the corporate network from and to the Internet.

On one side, the entrance of queries for the RTSP protocol (port 554), from Internet to a corporate IP address, has been enabled. The queries are forwarded to the RTSP server thanks to a NAT technique (Network Address Translation). On the other hand, enough bandwidth has been reserved for the output connection in order to assure a correct visualization of the VoD content streaming at the client side.

7.2.2.5 Users' set up: the receiver software

The hybrid *Netbox* receiver that is used in the test can give the user access to the web that lets the user choose, not only the Video On-Demand content that s/he desires to watch, but also the audio s/he desires to listen to: standard Catalan or audio description.

Figure 7.10 shows a screen shot of the selection menu that will be used by the user to select both the contents and audio channel:

Figure 7.10: VoD contents and audio selection menu

Note: For the test, we have used 4 episodes of a well-known Catalan sitcom produced by TVC, named "La Riera", and a film with AD, "Sense and sensibility".

7.2.3 AD enriched contents downloading for consumption on multimedia devices

7.2.3.1 Description of the scenario

This emerging scenario lets the user download files with media content on a PC, so that these contents can be visualized on the same PC or in another multimedia device. The following schema shows the distribution channel:

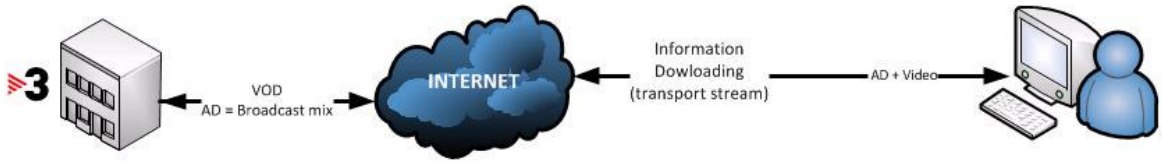


Figure 7.11: Distribution of AD enriched contents by downloading

The clips that will be available for the test will be exactly the same ones as in the previous scenario (streaming VoD). This way, users will be able to evaluate the quality of the contents in a comparative way.

During the test, the contents will be played on a laptop thanks to the VLC reproduction software[1].

7.2.3.2 Publishing and downloading system

The users will be able to download the contents to PC by accessing a web page with the links to the different programs that are all available with AD.

7.2.4 Audio description files downloading (Podcast)

7.2.4.1 Description of the scenario

This scenario lets a user download audio files in a portable format (MP3). These files only contain the audio description audio channel of the selected programs. In other words, these files contain AD podcasts. The users will be able to use these files in their portable audio player devices. One of the goals of the test is to discover whether there is a demand from end users to consume podcasts of the programs they have previously missed. For example, would users listen to the podcast of yesterday's sitcom they were not able to watch, in a car or on the bus while going to work?

The following schema shows the distribution channel:

[1] VLC media player is a player and multimedia framework of the VideoLAN project. It is open source software distributed under GPL license. More information at and downloadable from: VLC http://www.videolan.org/vlc/

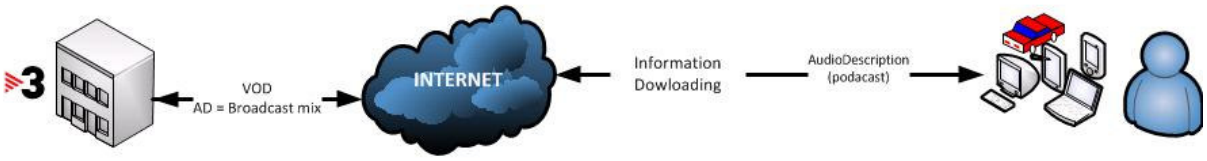


Figure 7.12: Podcast distribution

7.2.4.2 Publishing and downloading system

Users can download the audio description *podcasts* to a PC by accessing a web page with links to the different TV programs. Once the files are downloaded, they can be copied to the portable audio player devices, so that users can consume the contents wherever they want.

7.2.5 Simultaneous AD and non-AD consumption

The following scenario proposes a solution for family welfare (when watching TV) in those households with a member with visual impairment. It deals with simultaneous consumption of media content with audio description and without audio description.

The disabled person will be equipped with a laptop with DTT tuner, and will be able to listen to the audio description channel through headphones, while the other members of the family will hear the standard audio from the loudspeakers. However, the visually impaired person will watch the same image as the rest of the family, the one that is shown on the television screen. The aim is to ensure that this scenario is viable and that it does not involve audio synchronisation problems.

The scenario is sketched in the following scheme:

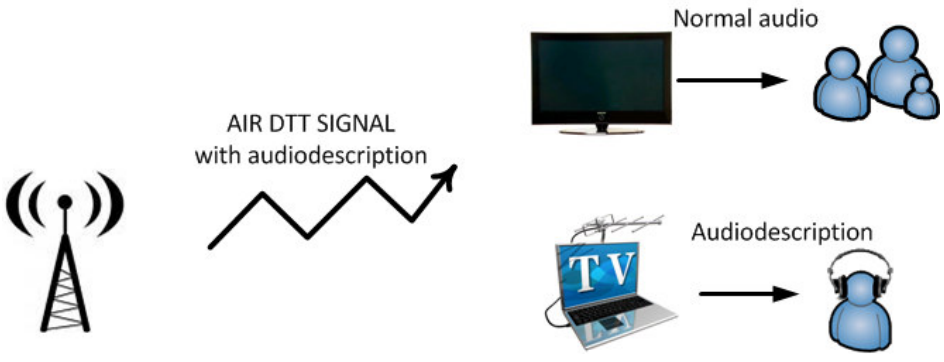


Figure 7.13: Simultaneous AD / non AD content consumption scenario

7.2.5.1 Users' setup

This scenario will use the standard DTT signal from TVC, Television of Catalonia, as it contains an audio description channel (broadcaster mix) besides the standard audio channel. The test set-up will include the following equipment:

- A TV with a standard DTT tuner. This equipment will be used to display the contents selected for the trials. Standard audio will be selected on this television.
- A laptop. We will connect to it a pair of headphones, and a DVB-T USB (Universal Serial Bus) receiver with DTT tuner: The laptop will tune to the same channel that is shown on the TV screen; however, the audio description channel will be chosen on the laptop. This way, the visually impaired person will be able to listen to the AD channel via the laptop, while s/he can sit with the rest of the family who are watching the TV screen.

8. Tests with a Demonstrator of an Enhanced Text Service

Partners involved: IRT, RBB

By Bettina Heidkamp

8.1 Introduction

The background to this demonstrator lies in the possibilities for enhanced access services enabled by hybrid (DVB-T/IP) receivers with broadband connections and the large screen diameters and high resolution displays of High definition Television (HDTV). Deliverable D3.3 gave a detailed description of this demonstrator, a barrierfree second generation digital TV-text service in the industry standard Hybrid Broadcast Broadband TV (HbbTV) that can be decoded and represented by HDTV receivers supporting HyperText Markup Language (HTML)/Extensible Markup Language (XML) using CE-HTML (Consumer Electronics-HTML). The barrierfree version of such a novel text service developed in cooperation by IRT and RBB was based on an existing service, the HbbTV text service for the German nationwide Channel 1 of ARD that had been launched at IFA 2009 (based on the existing teletext service for channel 1). This text had originally been conceived by the responsible department ARDText which is hosted by RBB for all ARD broadcasters. Among other ARD demonstrations this was actually the first public demonstration of an HbbTV service. This solution was developed jointly by IRT and ARD Text. It was decided to use this new service as a basis for demonstrating and testing a barrierfree novel text service in DTV4All. The idea was to tackle access issues of an HbbTV service based on CE-HTML and displaying mixed text, graphics and pictures. Sight impaired users were to be provided with the means to adapt the service to their personal needs and preferences for better legibility and understanding.

Please refer to Deliverable D3.4 for detailed information on the demonstrator and on all issues related to the framework of the user tests as well as to the three different questionnaires used. This deliverable D3.5 focuses merely on the results of these tests:

- D3.4 gives a detailed overview over the general approach, the concept and the implementation of this barrierfree prototype which was conceived in close cooperation of IRT, RBB and ARDText. Features were mainly a zoom view, different colour variants adapted to the requirements of different sight impairments and eye diseases and a Text-to-Speech (TTS) facility.

- On 14th and 15th December the service was tested with nine sight impaired people in individual user test sessions. D3.4 describes in great detail the methodological approach of this test, the set up of the room and the structure of each individual test session. It also elaborates on the methodological considerations as to recruiting a representative user group.

Summing up, RBB thought that the group should be representative of *the most important impairments/sight limitations* rather than of the different diseases as the origin of these sight limitations. Moving from there it was decided to try and find testers representing the following:

- Inability to see/distinguish red and green
- Inability to see colour, combined with limited vision
- limited contrast sensitivity
- visual impairment level 1 (WHO)
- visual impairment level 2 (WHO)
- visual impairment level 3 (WHO)
- blurred vision

With the help of the regional impairment associations we succeeded. The user group was distinguished by the different kind of eye diseases, by the different visual impairment levels according to the World Health Organization (WHO) and by different visual impairments as a result of eye diseases. A matrix displayed the characteristics of the user group is given in subsection 10.4.

8.2 General user profile as to media usage

Before the actual user trial, we interviewed the test persons to establish a user profile concerning their use of computers, television, and teletext. We wanted to find out if the visually impaired use the conventional teletext and furthermore, if those users who use computers and are familiar with the menu functions of web pages found it easier to use the HbbTV based ARD text.

All of the nine participants have a computer and eight of them use it daily, mainly for Internet access and emails. Office applications and audio/video come second and third. 22 percent of the participants also watch TV on their computer.

All of the testers have a television and most of them use it daily (89 %). The remaining 11 % use it at least several times per week. All of the participants are familiar with teletext. Most of them (33 %) use it several times per month, 22 % several times per week, 11 % once a week, 11 % don't use

it at all. When asked how satisfied they are with the conventional teletext font size, 33 % percent rated it a "three" and "five" respectively (on a scale from 1 = very good to six = very poor). The rest of them were equally divided, rating it "two", "four" and "six" respectively. None of the testers rated it as being "very good". When asked to rate the suitability of the colour in terms of legibility, the majority (56 %) rated it as being "satisfactory" (= i.e. a grade three). 22 % give it a "five" and 11 % "one" or "six". Approximately half of the testers are familiar with font size enlargement (56 %), 44 % were not familiar with this function. 44 % have already used it. For 22 % of the users the degree of enlargement is insufficient. 67 % of the participants use a digital set-top box.

Our test group is computer and TV literate. The testers use teletext at least sometimes despite their visual impairments, 22 % percent use it on a daily basis. One person uses neither a computer nor teletext. Correspondingly, this person had the most difficulties comprehending the navigation principles and obviously wasn't familiar with the "menu bar" concept.

Summarising the results regarding the legibility of conventional teletext: 55 % find it bad (rating it a grade 4 to 6), whereas 44 % think it is good or at least sufficient. What needs to be taken into consideration here is that some of the testers said they would use teletext if the font were larger.

8.3 Task Analysis

Fulfilling certain tasks, an indirect method for learning about the usability of the service, was the second main part of the test. The meaning and objective of each task was clearly defined. Concerning each "barrier-free feature" it was checked first whether the respective navigation for reaching the feature was user friendly and second if the solution as such was satisfying for the user by trying out the feature. The users' success or failure in performing a task thus provided evidence if the exemplary approach of the barrier-free features was convenient, user friendly, and most important useful for reaching satisfying access to the service. In addition, free spontaneous comments of the users were to show which features and functionalities they liked or disliked. The fulfilment of the tasks and the problems being encountered was measured by the usability team.

In the first task we presented the users with the ARD Text start page. We wanted to test if or rather how easily they would find the accessibility options.

Please activate the settings for the barrier-free version of ARD Text

	yes	with help	no
1.1 Comprehension and visibility of the "barrier-free" menu button for accessing the barrier-free version	56	33	11
1. 2 Comprehension of the arrow and OK key/s	78	11	11

User Rating	easy	moderately easy	difficult	too difficult
How easy was it to get to the barrier-free setting options?	56	11	33	0

The result shows that a small majority easily found, comprehended and saw the "barrier-free" button. One third, though, needed help. Only one person didn't manage the task at all. Accordingly, a small majority thought the task was "simple", whereas one third found it "difficult".

Compiled Observation and Testers' Comments

Three users initially had problems with orientation on the start page. Two of them first needed to make them selves familiar with navigation using the arrow keys on the remote control. Another two testers firmly criticised the positioning of the "barrier-free" button. They think the lower left-hand corner is not the best place for the "barrier-free" button. They argue you wouldn't expect the button there because the common reading direction is from top left to bottom right. With a restricted field of vision this would prove disadvantageous. "Top left is better. Starting at the top, it is strenuous to read until you get to barrier-free". (Note: A viewer with a restricted field of vision cannot comprehend the screen as a whole. It has to be read bit by bit, which is more laborious, obviously beginning at top-left). Both of those testers have scotoma or visual field loss.

One test person immediately tried to navigate using numeric keys. He was, however, unable to do this because the input window for numbers is not enlarged. Participants who found the task "difficult" gave the following reasons:

The positioning of the "barrier-free" button on the bottom left was again criticised. One tester was irritated by the television programme background. In her opinion the application should fill the whole screen. Even a still background image might not offer enough contrast.

Two testers thought the font on the start page is too small. The "barrier-free" button in particular is too small. Those participants have a visual impairment of level 2 or 3. Along the same lines, another person vaguely suggested "The start page should offer better accessibility." One tester said: "Arrow key navigation is too difficult for me."

Recommendations concerning the start page

- Barrier-free button should be enlarged and positioned top-left

- Input window for entering page numbers, i.e. for numerical navigation, in the barrier-free version should be enlarged

- Start page should be more accessible in general

- Application without background image (should cover the whole screen)

This task served to examine whether the page design for the barrier-free settings is sufficient.

	yes	with help	no
2.1 Comprehension and visibility of configuration page and setting options in general	33	56	11
2.2 Revision: Comprehension of the arrow and ok keys	89	0	11

User Rating	easy	moderately easy	difficult	too difficult
How easy was it to adjust settings according to your personal requirements?	56	33	0	11

The grades for managing this task are lower than for the first task, i.e. this task seems to be more difficult. Only one third of the testers succeeded without help, the majority needed support. Again, only one person (the same as before) didn't manage the task at all. Repetition, however, seems to have enhanced the comprehension of the arrow keys and navigation.

Interestingly enough, the self-assessment was more positive: A majority of 56 % said it was “easy”, only a third thought it was “difficult”. One person found the task “too difficult”. This wasn't the same person who seemed to have difficulties with navigation in general.

Compiled Observation and Testers' Comments:

Four testers chose the setting “white on blue”, another four users chose “white on black”, one of whom changed to “black on white” in task 5.

One test person had ongoing difficulties with navigation; another person had problems with using the remote control. She pressed “OK” twice, comparable to a double mouse click on a computer.

Three testers criticised the box ticking system as being unclear. None of them realised that by ticking the box they had made a choice and confirmed it. All three testers would like a confirmation signal, e.g. an audio signal. One tester chose “standard”, thinking that this would activate font enlargement, which he had selected beforehand by ticking the box. He thought he was setting font enlargement as “standard” in his personal settings.

The following reasons were given to explain why it was “difficult” or “too difficult”:

Two people expected an overall enlargement by choosing “font size enlargement”. This led to confusion because they didn't realise if they had really activated font size enlargement.

Four testers criticised the menu headers (headings) as indecipherable. Therefore, they could not fully understand the page. Those testers had a level 1, 2 or 3 visual impairment. Two of them specified: A header or headline has to be identifiable as such, perhaps defined by colour (yellow).

One tester would like all fonts on a page to have the same size because otherwise she would have to constantly modify her viewing distance. Three testers criticised the degree of font enlargement as insufficient. Those are testers with visual impairments of a level 2 or 3. One tester asked for the colour set "yellow on blue or vice versa". Another tester said: never use serifs. Another person would like a clearer focus (noticeable frame).

Summary: Recommendations – Preferences Page

- Confirmation signal when box is being ticked (audio signal: "You have activated font size enlargement")
- Reconsider the term "standard", it seems unclear
- Reconsider the term "font size enlargement", it also seems unclear. Maybe use "zoom function" or something similar instead.
- Use same font size for the entire page.
- The headings (headlines) are indecipherable
- The headings (headlines) should be clearly identifiable as such (colour?)
- Increase the degree of font enlargement
- Colour sets "yellow on blue" and "blue on yellow" are requested
- The frame (focus) should be clearer

Please go to business news page

	yes	with help	no	didn't use it
3.1 Comprehension of navigation principles (from centre to menu bar and back)	89	11	0	0
3.2 Legibility of the menu bar	100	0	0	0
3.3 Where applicable: Comprehension / Visibility of the numerical navigation	33	0	22	44

User Rating	easy	moderately easy	difficult	too difficult
How easy was it to find the business news page?	89	11	0	0

Compiled Observation and Testers' Comments:

Navigation was understood by 89 %. As before, this is a satisfactory result.

One person had ongoing problems. A positive result was that all of the testers thought that the menu bar (with font enlargement) is legible, and all of them thought this task was “easy” or “moderately easy”. Nobody used numerical navigation.

In explaining why the task was easy or difficult, one tester remarked it confused him that there was no space between the page number and the respective header. Three testers criticised that the cursor (focus) jumps back to its original position when switching from the centre of the screen to the menu bar – i.e. the cursor then is out of view. They don't see it any more and loose orientation. Those testers would like the cursor to stay at its current height while switching between different fields on screen.

Summary: Recommendations – Locate News

- When switching between different fields on the screen, the height of focus should the same.
- Distance between the page number and the respective header is too small.

Please read news report “Quelle: 'White Goods' Sale“

	yes	with help	No
4.1 Comprehension of scrolling principle, scrolling from news headline to news headline	89	11	0
4.2 Legibility of news headlines with the contrast selected	100	0	0
4.3 Comprehension of the “OK” or page numbers for selecting news item	89	11	0
4.4 Legibility of news report with the contrast selected	100	0	0

User Rating	Good	sufficient	bad	not at all
Please rate: How legible was the news text for you?	89	0	11	0

The result is very positive: 100 % could read the **news headline** on the overview page as well as the **news text** out loud, though, sometimes haltingly as is to be expected.

Almost all of them (89 %) understood the scrolling principle, **scrolling the news headlines** and selecting a report by pressing "ok".

Compiled Observation and Testers' Comments:

One tester criticised that he could only read the news report when sitting very close to his TV (approx. 8 cm distance: visual impairment level 3). He stated explicitly: "The font enlargement is insufficient." Another tester thought that legibility is reduced compared to enlarged font in standard teletext. By contrast, two other testers remarked on the good legibility. One of them even hoped he will now be able to read the content without a magnifying glass. Both testers have a visual impairment level 2.

One test person said the news headline page is confusing and suggested a clearer structure (maybe by using a yellow font). She and two other testers would like a "back" function.

Two of the testers faced the problem that while scrolling down to read a report they suddenly were in the next news report (a task which is to follow next). One participant states, she liked the scrolling principle for reading in general. Nevertheless, it obstructed and confused her that by scrolling down another report follows immediately. She found it especially annoying that the content "jumps by sections" as soon as a new report appears. The other tester said: "I didn't comprehend that another report follows immediately".

Finally, three of the testers said that the cross pad on the remote control should better be separated as it is "too confusing" to have all the functions on one button.

Summary: Recommendations – News Overview Page and Report Level

- For users with a visual impairment of level 3 the font size seems to be only partially sufficient, especially for continuous text.

- There is demand for a “back” function.
- It adds to confusion that in scrolling through a news report, the next report is immediately opened and one can also move to the next news item by scrolling in sections.

Please read the next news report (Mechanical Engineer ...)

	yes	with help	no
5.1 Comprehension of browsing principle to move from one news report to the next, in combination with scrolling through a single report	22	56	22

User Rating	easy	moderately easy	difficult	too difficult
How easy was it to go to the next news report?	89	0	0	11

As already described in task 5 this point proves to be problematic, despite the very positive self-assessment: 89 % found the task “easy”. But this is due to the fact that most testers use the menu bar or the overview page to open a report. They don’t move from one report to the next. Thus, the majority (56 %) needed help to go directly from one report to the next.

Compiled Observation and Testers' Comments:

Evidently, five people didn't fully comprehend the principle that the next news report follows directly. Four users noticed it while reading. One person, nevertheless, went to the overview page first.

One person still has difficulties comprehending navigation and scrolling.

Commenting on their “user rating”, seven people state that they don't approve of the system scrolling down to the next news report, they prefer the function to be more distinct or accessible via a special button. Two users suggest a visual impulse or audio signal when scrolling to the next news item. One person would like to scroll through the report line-by-line instead of “report-by-report”.

Again, four people would like a "back" function after reading a message.

Summary: Recommendations – Browsing from Report to Report in Combination with Scrolling (also see previous task)

- The system "scrolling from report to report" should be reconsidered
- If this function is kept, it should contain:
  - Visual or colour impulse or audio signal
  - Line-by-line scrolling
- Introducing a "back" function to move back to the overview page

Please choose setting "ARD Text voice output".

This task is a revision of the preferences page evaluation.

	yes	with help	no
6.1 Revision: Comprehension and visibility of navigation (opening page "barrier-free settings").	78	11	11
6.2 Revision: Comprehension and visibility of the preferences page and setting options in general	89	0	11
6.3 Revision: Comprehension, visibility and selection of specific setting options	89	11	0

User Rating	easy	moderately easy	difficult	too difficult
How easy was it for you to select option „ARD Text voice output"?	89	0	11	0

As a positive result, the task "comprehension and visibility of preferences page and setting options in general" achieved 89 %. This is much higher than at the beginning where only 33 % didn't need help with the page. Accordingly, the testers' self-assessment is slightly more positive, 33 % said the task was "easy". This may suggest that the testers have by now become more familiar with the Text. It is important also to acknowledge that this page was explained to the testers during task two. It

seems slightly more difficult to access the preferences page via the start page. One tester needed help in doing so; another tester didn't manage it at all. The testers' self-assessment is positive: all except one say it was “easy”.

Compiled Observation and Testers' Comments:

Seven of the testers think the task is “easy”; one person forgot to press “ok” and thought it was sufficient to click on the “barrier-free” button. One person criticises that she can't see the preferences page before the font size enlargement has been activated. One tester suggests using another, in his opinion, better synthetic voice (“Julia” and “Klaus”).

Summary: Recommendations – Preferences Page

- Voice is okay but there are better ones
- Larger font size already on the preferences page (even before font enlargement)

Please go to news page

This task aimed to examine if and how the testers use the now activated voice output to access the news overview page (comprehension of the voice output for first level navigation). Here, the testers could only access the news page via the numbers read by the voice output (or from memory alone, if they listened to the voice output of the menu bar from beginning to end). We also wanted to see how well the testers could hear the voice output.

	yes	with help	no
7.1 Audibility and comprehension of voice output for navigation on the menu level	0	22	78

User Rating	yes	no
Was the voice output helpful to open the news page?	11	89

In quantitative terms this is the worst result thus far. The vast majority of the testers didn't understand the **voice output for navigation on the menu level**. In their self-assessment all but one tester said that the voice output wasn't helpful to them.

Compiled Observation and Testers' Comments:

The difficulties were not an **acoustic problem** as the testers understood the synthetic voice well. The voice output explicitly confused four of the testers; they prefer to use visual navigation principles as usual. Another tester said that he concentrated on the screen visually and therefore wouldn't use voice output.

Commenting on their self-assessment, eight of the nine testers said they understood but ignored the voice. One of them said “the voice is annoying”. Five people criticised that the whole menu is being read out to them. They think, only what is in focus should be read by voice output. All standard screen readers used by the visually impaired testers are structured that way.

Summary: Recommendations - (Menu Level/Navigation)

- Voice output for navigation doesn't appear to be absolutely necessary.
- In its present form it is insufficient and confusing
- Voice output appears useful only if it reads what is in focus.

Please select voice output for the news report “States demand …“

This task focuses on two aspects: Using voice output to find a specific news report on the overview page level (the news overview page is read from beginning to end, the cursor is initially positioned on the first headline in the overview). Again, users don't have the choice here to open a report just while the headline is being read out (no link of voice output to chosen focus). They can select a news report only by entering the page numbers that are being read out.

	yes	with help	no
8.1 Audibility and comprehension of voice output for navigation on news overview page level	11	11	78
8.2 Audibility of voice output for news report	100	0	0

User Rating	yes	no
Was the voice output helpful for finding a particular report?	11	89

Compiled Observation and Testers' Comments:

Again, audibility of the content was perfect (100 %). On the other hand 78 % did not comprehend the navigation. Only one person understood the principle. Another tester understood it after receiving some help. Six of the testers didn't use voice output at all or only marginally. One of them explicitly stated that he read the page to find the news report.

Two of the testers criticised that they could not select the headline through the voice output. They would like the headline to be a focus which can be activated. One tester said the voice was "pleasant", even over a longer period of time. Another tester thought it was a good idea to combine reading with listening. She would have liked the overview to be read out twice. Another tester also favours a "mixed use". Voice output is helpful for longer continuous text but for navigation it is too difficult.

Again, three of the participants suggest that voice output should only read what is in focus. Two testers said voice output is confusing if the user can still read relatively well.

	well	sufficient	not well	not at all
How well did you understand the report?	56	44	0	0

The result for audibility of the report is very good. Five of the testers said explicitly, they understood the voice "well". But one of them thought, the voice sounded a little muffled.

Summary: Recommendations – Voice output of News Overview Page

- Voice output without link to focus confuses the users
- Mixed use of "reading – listening" could be helpful, voice output only for continuous text

8.4 Assessment of Evaluation Interview (Past task interview)

After having fulfilled the tasks, the users were asked direct questions to find out about wanted qualities and functionalities after the fulfilling of the tasks was finished.

1. General feedback on the barrier-free version of ARD Text

a) What do you like best about barrier-free ARD Text?

	%
Font enlargement	78
Colour selection / font convertibility	78
Individual settings	56
Computer Based Handling (familiar)	33
Voice output	33
Clear menu structure	22
Easy adjustment	11
Easy to operate using arrow keys	11

At this point, the testers freely expressed their opinions. Their spontaneous comments were categorised in order to compile the main points and get quantitative results. Of course, most of the users noted several points. The two functions **font size enlargement** and **colour selection** clearly are on top of the list. The testers were delighted with these functions.

These results were confirmed by the fact that more than half of the testers like the **individual settings**, which mainly offer font enlargement and contrast settings. The comments show that the basic features for accessibility are successfully implemented. Only three of the testers mention the voice output, generally it receives a better rating than would have been expected according to the test results; it is explicitly mentioned as the best feature. The remaining three features, though not mentioned as often, altogether imply that the text proves to be user-friendly and clear.

b) What do you like least about barrier-free ARD Text?

	%
Missing More and higher degrees of adjustment are missing. The font size, even when using zoom, seems smaller than in standard teletext.	44
Voice output should only read what is in focus (as all standard screen readers do; the usual standard should be followed)	22
If I don't see well, I can't even find the page with the accessibility options. The "barrier-free" button is not visible, it's not positioned correctly (top left would be better), suggestion: a configuration page where you can choose between "normal" and "accessible version".	22
Perhaps it would be better to enlarge all content on the screen?	11
The clock should be bigger (positioned above the programme tips)	11
Cursor should stay at one level during navigation to the right or left, otherwise it can disappear from the field of vision and has to be searched for	11

Again, there is a positive tendency in the evaluation, which confirms the basic concept, as the testers point out more positive than critical features. 44 % of the testers wished they could enlarge content even more. This is made explicitly clear with the critical comment: "The text seems smaller than normal teletext, even when using the magnifying function". On the same lines, one person would like a general enlargement of the page/screen instead of having a magnifying function.

Twenty percent vote for a voice output which is linked to the focus. Twenty percent also criticise the access to the configuration page. They think the barrier-free button is not only positioned wrongly but also difficult to see. Other critical points were the size and positioning of the clock, and the jumping of the cursor/focus.

All the critical points that were explicitly mentioned in these open comments had also proved difficult in the task section (and are therefore marked red and bold below in the final resume).

Key questions

a) Would you be interested in this kind of new text service on your TV if it were barrier-free?

	%
Yes	100
No	0
Comments / Other:	
I'd rather use it on the computer	

These results again underline the essentially positive trend in the evaluation of the general concept.

b) There are different ways to offer a new kind of TV teletext service. Which do you like the best?

	%
1. Adapting the standard text service with an accessible version (as evaluated here).	78
2. A graphically reduced, text-only version, which is made for and offered to the visually impaired.	22
Comments / Other:	
Whatever is the easiest to use	11
A reduced version like that mentioned in option 2 could lead to "disimproved" handling; omitting a lot could lead to complications, a general version which is accessible for all is best	11

A majority of the testers opt for an accessible version of the existing text for the visually impaired. Again, this confirms the general concept.

Accessibility Options

Font Size

a) How sufficient was the degree of font size enlargement for you

	%
appropriate / sufficient	78
too small	22
Too Big	0
Comments / Other:	
Maybe offer another two or more levels of font enlargement	33

A clear majority of the testers think the offered font size enlargement is sufficient. This is surprising, considering the task results above and at point 1b (open assessment). However, after having answered the multiple choice questions, a third of the testers tackle this point on their own initiative, requesting a higher degree of enlargement.

b) How important is it for you to be able to adjust colours and contrasts?

	%
very important	89
important	11
not important	0

The ratings confirm the positive trend in the evaluation of this feature. It definitely should be kept.

Voice output

a) How important is voice output to you?

	%
not important	56
Important	33
very important	22
Comments / Other:	
Helpful for certain features (long continuous text)	11

Effectively the same percentage of testers consider this feature to be important/very important as not important. The task results show that this feature is more controversial than colour and contrast settings. On the other hand, there is definitely a demand for it, especially in regard to longer continuous text.

b) How important is a speed setting for voice output to you?

	%
not important	44
Important	33
very important	22
Comments / Other:	
Would be a nice feature	11
Important for information	11

A speed setting function is important/very important to a majority (55 %). To 44 % this feature is not important. 22 % of the testers confirm their multiple choice selection with comments.

c) Do you have further comments regarding the voice output?

	%
voice okay	33
voice when in focus	33
choice of different voices (perhaps female and male)	22
swallows syllables / clearer diction would be better	22
synthetic voice / computerized voice	22
on-off switch for selective use of voice output	22
voice output is very helpful for longer continuous text	11
screen reader should incorporate standard functions	11

The free evaluations vary to a great degree. One third of the testers say they like the voice. That is slightly contradicted by the fact that 22 % of the testers criticise the voice, saying it sounded synthetic, it failed to pronounce some syllables and they would like the option of having a male voice. In summary, the voice was good on the whole, but should be improved (which is possible without much effort) if a version with voice output is to be offered. A further third of the testers readdressed the point concerning the voice output function with respect to focus. Here, the testers also voice a desire to stick to the usual screen reader standard. In addition, several participants express the wish for voice output specifically for news reports.

Contrast / Colour Sets

a) How important is it for you to be able to select colours and contrasts?

	%
very important	67
Important	33
not important	0

The result is positive; the ratings are comparable to the results regarding the enlargement function. This feature should definitely be kept.

b) Which sets of contrast variations do you need for the accessible version?

	%		
	very important	important	not important
White on Black	44	33	22
Black on White	11	44	44
Yellow on Black	0	33	67
Green on Black	0	33	67
Blue on White	0	11	89
Yellow on Purple	0	22	78
Yellow on Blue	0	44	56
White on Purple	0	11	89
White on Blue	11	67	22
Yellow on Red	0	11	89
Green on Blue	0	0	100
Yellow on Green	0	11	89
Green on Purple	0	11	89
Green on White	0	22	78
Green on Red	0	0	100
Red on Black	0	0	100
White on Green	0	22	78
White on Green	0	33	67
Blue on Green	0	0	100

1
2
3

Conclusions:

Colour sets included in the test version which should be kept:

- White on Black
- Black on White
- White on Blue

Additional colour sets which were not included in the test version:

- Yellow on Black
- Yellow on Blue

- Green on Black
- White on Red
- Blue on Yellow (derived from tasks)

Colour sets which can be eliminated from test version:

- Blue on White

Font Size

a) How important is it for you to be able to choose from different fonts?

	%
not important	78
important	22
very important	0

This feature doesn't appear to be important. The testers, however, state that a font without serifs should be used.

Questions concerning the design of colour buttons

Colour buttons will play a large role in internet-like services for television. These buttons will be used to activate television services such as the teletext service you tested today. We need your opinions to design the best possible barrier-free buttons. Here are a few suggestions.

	%		
	very good	good	bad
Colour buttons (FT) on remote control and screen are distinguished by form and colour	11	11	11
Screen buttons are distinguished by a), b), c), d)	0	0	11
Screen buttons are distinguished by FT1, FT2, FT3, FT4	0	0	22

None of the solutions we suggested received much approval among the testers. The best solution, which only received 22 % approval, is to use form and colour to differentiate between the buttons (a form-icon on the screen in addition to text on the button. Stamped or punched form on the remote control, as print often rubs off quickly). Users with complete colour blindness would have no other way of differentiating the buttons, and users who can't see well could also recognise the different forms. But this would also necessitate a change in the design of the remote control.

Comments / Other:	%
Colour buttons are indecipherable	78
Higher Contrast (Font background)	44
Brighter colours without gradient	33
Basic functions should be the same as in standard teletext. New menu functions better on extra menu bar	11
Has to be a barrier-free version	11

The key factor: In the free commentary, the majority of testers make a point of mentioning that they couldn't read the colour buttons at all, because of the lack of contrast and the pale colours. Because of that, 44 % of testers say there should be more contrast between the background and the text. One third also wanted flashier, brighter, colours without gradient. Even stronger colours like white on yellow would be unreadable. Better combinations would be black on yellow, white on bright red, white on dark blue or white on dark green. The colour buttons and the text written in them must be bigger, and should also be magnified when using the magnifying function. These changes would not be of any use to those with colour blindness. In their case, the form of the buttons must vary (see above).

Do you have any other comments regarding the barrier-free design of this new television text service?

	%
Delighted	*33*
A general configuration page for "standard" and "barrier-free version" is requested	22
Input window for entering page numbers should be larger, you should be able to directly control it, current position must always be clear	22
Separate keys for colour set selection or inversion. Screen readers do have a key for inversion and a second key for colours, that should be implemented here	22
Operable	11
Navigation using the cross pad easily learnable	11
Much improved compared to teletext	11
Headlines of categories are much too small	11
Higher level of continuously variable magnification	11
Would like to use colour buttons	11
Comfortable computer-like navigation	11
Spacing between characters should be larger and/or you should be able to select spacing and font-weight, sans-serif fonts are generally better	11
Should have a function to also magnify graphics	11
More feedback: e.g. if zoom function is activated	11
Programme should be enlarged	11
Line breaks aren't clearly defined	11

One positive observation here was that the majority of the entries once again express enthusiasm for the text service: One third of all testers expressed their approval. Three other entries gave positive feedback (all in italics). The red highlighted points are comments that were repeatedly mentioned (in the tasks and this evaluation section) and should therefore be taken into account.

8.5 Results and conclusions summed up

The new barrier-free equipped HbbTV-ARDText was very well received by visually impaired testers. Some testers even expressed a great deal of enthusiasm about the text service. Many of the testers found it to be a real improvement compared to traditional teletext. Testers found the customisable font size enlargement and colour/contrast settings especially helpful. They liked the "computer-like handling," and thought the structure was straight forward, on the whole.

The user trial did expose a few "weak points" to be considered if this text service should be implemented. The users were extremely engaged and constructive. They made many suggestions for improvement. These included many points that were not obvious to us as sighted individuals, and therefore weren't part of our concept of a barrier-free version.

This section offers a detailed résumé of the task accomplishment and free commentary sections. All points marked **bold** were mentioned both during the user trial and the interview (mostly in the free commentary, sometimes in the multiple choice questions), therefore they carry more weight, relatively speaking. These will each be explained in red. A person with level 2 visual impairment observed the entire test and made the comments in blue. This person also later sent us valuable comments of her own accord, even though she did not participate in the test herself. Nevertheless, these comments should be included in the results.

Suggestions - Start page:

- **Barrier-free button should larger, and in the upper left-hand corner.** This critique was mentioned several times during the user trial, and 22 % of users mentioned it during the "What do you like least" interview.
- **The input window for page numbers should be enlarged in the barrier-free version.** This critique was also mentioned during the user trial and again, from 22 % of users, during the "closing remarks regarding the barrier-free design of this new television teletext" interview category.
- **The configuration page should be "more barrier-free" as a whole. Obvious from** task results, and based on 22 % result in interview category "what do you like least."
- Application should not have a background image (should cover the entire screen).
- A separate barrier-free entry page with the choice between "normal" and "barrier-free." (22 % of users mentioned this during free answers in two different interview categories).
- The clock should be larger and above the programme tips (interview: What do you like the least?)

Suggestions – Configuration Page:

- **Confirmation of ticked boxes (e.g. sound signal: "You have chosen the degree of font size enlargement.")** This critique was mentioned during the user trial and in the interview ("more confirmations" – 11 % in the "closing comments regarding design …")
- Rethink the term "standard." It appears to be confusing.
- Rethink the term "font size enlargement." Maybe "magnifying function" is better.
- All fonts on the page should be the same size.
- **The headings are unreadable (too small).** This was mentioned during the user trial, and one user mentioned it again in the interview ("Headlines are too small." Free commentary category "Closing comments …")
- **The headings should be more obvious as such (colour?).**
- **The degree of font size enlargement should be higher.** With 44 %, this was the most serious negative point in the free commentary "What do you like the least?" In the multiple choice questions on font size enlargement, although 78 % said it was satisfactory, one third commented voluntarily "Maybe two sizes bigger or more." In the category "Closing comments" one tester expressed a desire for "more continuously adjustable enlargement." The commentary made by the accompanying person who sent an evaluation: "The font size should have several adjustable levels (normal, large, very large)."
- Additional points on the **theme of enlargement**:
    - **You should be able to access the input window for entering numbers directly and it should be larger. It should also be clear where you are. (User trial and 22 % of users in free commentary "Closing comments regarding design …).**
    - In the interview "What do you like the least," 11 % expressed the desire for larger text as a whole (free commentary).
    - The clock should be larger (free commentary, 11 %, Interview "What do you like the least?).
    - The programme preview should be larger (free commentary, 11 %, Interview "Closing comments regarding design…")
    - Size of graphics should be customizable. Free commentary, 11 %, Interview: "Closing comments regarding design of barrier-free…. ")
    - **Users desired colour sets "yellow on blue" and "blue on yellow".** The desire for "yellow on blue" was confirmed during the interview (desired variations). Users were not asked again about "blue on yellow."
    - The frame (focus) should be easier to see.

- "The page number next to the search terms makes sense and is good. Perhaps all page numbers should be listed for entries with several pages (TV programme 300-308). That would make navigating with numeric keys easier."

Suggestions – Configuration Page (in the later repetition):

- **The text should already be larger on the configuration page (without needing to enlarge font first).**

Suggestions – Locate News:

- **When switching fields, the focus should always remain on the same level.** This was mentioned both during the user trial (mentioned by several users) and in the free commentary "What did you like the least?" (11 %).
- The space between the numbers and the corresponding headline is too small.

Suggestions – News Overview and Report Page:

- **For users with visual impairment level 3, the font size doesn't seem to be large enough, especially when using continuous text.** This observation confirms the above assessment under "Suggestions - Configuration Page" ("There should be a wider degree of font size enlargement.")
- Users desire a "back function." Mentioned by several testers during task accomplishment. (see below)
- While scrolling and reading a news report, and also with scrolling in sections, the next news report is immediately opened. This is confusing.

Suggestions – Browsing Principle between News Reports in Combination with Scrolling

- The system scrolling between news reports should be reconsidered.
- If this function is kept, it should contain:
  - Visual/colour impulses or an audio signal
  - Line-by-line scrolling
- Users request a "back button" which takes them back to the report overview page.

Suggestions – Voice output

- General needs:
  - In the interview "How important is voice output to you?", just as many users thought the function was very important/important as those who thought it was not important. In the "What do you like the best?" category (free commentary), though the **voice output** was only mentioned by 33 %, it was rated better than expected, judging by task accomplishment. Conclusion: Voice output is not as important as font enlargement and contrast, which were very clearly desired and praised. Nevertheless, a third of the testers listed it under the best text functions. The accompanying person stated: **"Voice output is very good and very important for all types of vision impairment."** (This sentence in bold text was the only one that this person wrote.)
- Design:
  - The current form doesn't make sense. It only confuses.
  - **Voice output only makes sense if it only reads what is in focus. Voice output not limited to what is in focus confuses the user.** This was a very obvious critique during the user trial, and was mentioned by 22 % of users in the free commentary interview "What did you like the least?" In the free commentary interview "Comments regarding voice output" it was mentioned by 33 %. It was indirectly mentioned by 11 % when discussing the topic "Following screen reader standards."
  - Voice output doesn't seem to be necessary for navigation.
  - **A kind of mixed use "read/listen" function could be useful. Voice output only for continuous text.** This was revealed during the user trial and was also mentioned by 33 % during free commentary interview "comments regarding voice output." One user also mentioned this during the free commentary interview (Other comments on the theme "How important is voice output to you?").
  - 22 % (free comments "comments regarding voice output") expressed the desire for voice output that can be turned on and off at will.
  - A speed setting was very important/important for 55 % of users, and not important for 44 %.
- Voice:
  - It seems the voice was understood well. One third of the testers found it pleasant, while others stated it is "all right." If voice output is going to be offered, a better voice should be chosen. Testers recommended the voices "Klaus" and "Julia."

Suggestions – Contrast Variations:

Variations from the test version that should be kept:

- White on Black
- Black on White
- White on Blue

Additional variations that should be offered:

- Yellow on Black
- Yellow on Blue
- Green on Red
- White on Red
- Blue on Yellow

Variations that can be removed from the test version:

- Blue on White
- Graphics (weather, for example) should contain a greater amount of contrast.

Suggestions - Font:

- This feature does not seem to be important to the testers. The deciding factor that is crucial for the testers is that a sans-serif font is used. Although in the free commentary to "closing comments regarding design," one user expresses the desire for "settings to control the spacing between characters and the font-weight." One further criticism: "the line breaks are not clear."

Suggestions – Design of the Colour Buttons:

In their current design, the colour buttons are not legible for 78 % of users. (Due to colour gradients, weak contrast because colours are too light and text is too small.) The colour buttons are obviously of no use to colour blind users.

Suggestions were:

- Higher contrast between the background and text (black on strong, "dark" yellow, white on strong, dark red, white on dark blue and white on dark green).
- Very bright, strong colours.

- No colour gradients.
- Larger text and/or enlargement possibilities.
- Differentiate the colour buttons using form as well (a form-icon on the screen in addition to text on the button. Stamped or punched form on the remote control, because print rubs off easily).
- The colour bars on the bottom of the screen should have more contrast. The size of the bar depends on the text content. The text content should be easy to read and easy to reach using the navigation buttons (so the font size can also be adjusted individually.)
- "I'd like to use colour buttons" (free commentary, 11 %, interview: "Closing comments regarding design…").

One more suggestion regarding barrier-free design of remote controls in general:

- "The industry should finally create an easy-to-use, universal remote control for all conventional devices. This would finally put an end to having to use multiple remote controls. The buttons on the remote control should be as large as possible, but there should also be as few of them as possible. The buttons should also have a night function similar to that of cell phones, where the buttons light up for about 5-10 seconds when in use.

9. Tests with a Demonstrator for Text to Speech Applications

Partners involved: RBB, IRT, AMMEC

By Bettina Heidkamp

9.1 Introduction

All methodological aspects as well as a detailed description of the set-up of the user tests and the composition of the user group are contained in deliverable D3.4. This deliverable D3.5 only documents the **results** of the test. However, this short introduction is to sum up the test's framework and aims for better understanding. Tests with the German speaking Ammec-device, a Multimedia Center for DVB-S, T or Cable which provides audio interfaces, were carried out at RBB in June and July 2009 with blind test persons. The envisaged laboratory test focused not so much on validating the Ammec device as such. Instead, it focused on generally validating "Audio User Interfaces" (AUI) for TV sets / digital receivers provided with the help of a text-to-speech engine (TTS). The Ammec was used for the test because at the moment it is the only TTS-based set top box on the German market. From the concrete, Ammec-related results, general conclusions concerning this topic were derived.

The laboratory test had two main aims:

General aim:

- To find out the *general acceptability* of such a service: Are spoken or audio interfaces generally welcomed by the target group, do "they make sense to the users" in terms of achieving improved access to digital television?

Specific aim:

- To find out how such a service should be designed in order to meet the requirements of the target users mainly in terms of the scope of functionalities but also in terms of basic navigation issues.

The individual user test sessions, which lasted about one and a half hours per test person, were based both on direct and indirect methods. To gather the information needed to answer the question posed in the *general aim* a direct method was chosen. The testers were asked respective questions by the interviewer after having tested the Ammec and having fulfilled certain tasks.

To gather the information needed to answer the *specific objectives* the method employed was both indirect and direct. Indirect: One important part of each individual user session was fulfilling various tasks referring to handling the Ammec. The meaning and objective of each task was clearly defined. The users' success or failure in performing a task provided evidence if whether the exemplary technological approach of the Ammec is convenient and user friendly in order to reach generic recommendations for audio interfaces. Furthermore, free spontaneous comments of the users showed which features and functionalities they liked or disliked. Fulfilment of the tasks (success) and the problems being encountered was measured by the usability team.

The detailed results of the task evaluation and of the final question and comment part of each individual user test session are contained in section 10.1. This following chapter summarises the important conclusions.

9.2 User profile

The tests were conducted by RBB with the help of nine blind users and one sight impaired user.
The users were recruited with the assistance of the Associations of the Blind in Berlin and Brandenburg. With one exception, a sight impaired woman, only fully blind testers were chosen in order to have at least a similar starting point for each tester. However, there was heterogeneity in the group. There were five persons who were born blind or went blind early in childhood, two persons who went blind in later childhood or younger adulthood and two persons who went blind late in life. It was obvious that the latter two persons were not familiar with text to speech technology and electronic support tools and dealt differently with the device than those who went blind early in life. The test group encompassed six men and four women.

Our "Pre-task" questionnaire opening each individual user session in order to see how computer or "media literate" the candidates were, showed that all 10 users were possessing a computer and were using it daily. While all used it for emailing and Internet access eight testers had also used it for audio and video services. Of these, seven used it for CDs and DVDs, four each for podcasts and TV Media Archives and three for receiving TV programmes. All testers had a TV set, six used it daily, one never. All testers knew what teletext is, 50 percent had used it, and the other half had not. Those who had used it relied on special cards for PCs (two testers), a screen reader for PC (one tester), or somebody reading it out to them (two testers). Six of the testers had a digital TV set top box and 8 had experience with recording media. All ten testers were radio listeners, all used traditional radios, seven, in addition, the PC, and four their set top box.

9.3 Summary of test results and conclusions

The test results show very clearly that a TTS-based device (Set-Top-Box) providing a wide scope of functionalities is very much desired by the testers. All our testers would use it and find it very important. All the tested functionality areas were deemed “very important” or “important” by our testers. This includes, apart from choosing TV channels, the Electronic Programme Guide (EPG) and Radio-related functionalities as well as recording and cutting TV and radio programmes, accessing read-out Teletext services and, a little bit less important, playing CDs and DVDs.

9.3.1 Operational concept

A very important general point for introducing TTS-based devices for digital TV is the need to have a very clear and logically structured operational concept. This is especially important for blind users to improve orientation. In the case of the specific device tested by us, the Ammec, the operational concept is structured into three major areas. These three areas apply to each main section of the menu. The areas are: 1. the title of the menu section, 2. the specific use of colour buttons for this menu section and, 3. the individual functionalities of this menu section. If users have internalised these three areas, orientation should be granted. However, in the course of our test we had to learn that a lot of misunderstandings occurred and orientation was not optimal. Obviously, the users had not been able to internalise this operational concept in the quite short time we had given them for preparing for the tests. The users had had an introduction / presentation of the device by a blind person (about 1.5 hours) and they had been provided with the manual and a remote control to take home. They were advised to study the manual and remote control for about two hours. This preparation was obvious not sufficient. The detailed test results show that the testers often asked for functionalities which were provided by the Ammec but which they had not found. Our observations showed that most users had problems in understanding the structure of the menu. A more complex approach could be the development of a small learning programme which would be automatically accessed on first usage. This leads to the conclusion that not only a good operational concept is of utmost importance but also profound and thorough instructions concerning the concept. An adequate introduction should be based on detailed (audio) manuals and probably additional tutorials. It might also help here to provide the possibility of accessing context-sensitive help pages.

Direct buttons

Users would like to be able to control certain functionalities directly by means of the remote control. They referred to the following functionalities in this context:

- Turning up and down the volume both of the TV sound and the audio interface
- Separate mute buttons for the audio interface on the one hand and the TV sound on the other hand
- Direct entering of numbers
- Choosing the next and the preceding channel
- "Read out"-button for the current channel for better orientation
- Colour button for accessing the EPG
- Unlock key for all their inputs

The Ammec device offers all listed points based on the operational concept which relies on different colour button functionalities for the different menu sections.

Concerning the mute button users said such a mute button should be easily recognisable and clearly separate from the other buttons. They would like such a button in case of something happening outside TV usage or whenever the user wants to continue listening to the TV programme and then later return to the videotext. The Ammec approach is a button which interrupts / mutes the audio interface. When pressed once again, this button re-activates the audio interface at the currently selected point. Users furthermore suggested a corresponding fault protection function, like, for example, a "a warning"-signal whenever further buttons are being activated while the mute mode is activated (similar to the signal/function known from password entering while the caps lock key is activated).

Less important for the users is a direct remote control button for accessing videotext and an "on - off" button. Accessing audio description services directly via the remote control was not part of the test, is, however it is definitely a very important functionality for blind users.

9.3.2 Prominent areas of interest

1. Functionalities

The testers asked for the possibility of individually influencing the volume and speed of the audio interface (independent of the volume of the TV-sound). Ammec provides this function.

In order to find their way inside videotext services, users would like to have read out page numbers while they are actually entering them. This is understandable as blind users have no feedback on

their inputs otherwise, which could lead to confusion with respect to page navigation. Testers were also interested in having text or tables etc. read out as single words or even letters. The Ammec device offers line-wise reading out. This function is especially relevant for videotext usage and also in the case of very long programme information text passages in the EPG.

Users missed an audio function for the time announcement, even though the Ammec device offers this (operational concept, area one "1. Title/first line being read out") However, as explained above, such misunderstandings occurred quite often as the users had not sufficiently internalised the operational concept.

Users would like to determine the sequence of channels themselves and even to create folders and subfolders along their interests or different service providers. Ammec offers sorting and editing of channels via colour buttons. A possible option could be separate buttons for TV and radio starting each with "One" for moving up and down the channels. This approach, however, could be regarded as opposing the idea of inclusion as it would differ from current set-top-box conventions.

Users opted for choosing profiles like "beginner" or "advanced user". Furthermore some users criticised the artificial voice. This was especially true for users who became blind later in life. Especially for such users a more "human" voice would mean a real improvement.

Concerning the device Ammec as such, users criticised the price, slow reaction times and that it does not provide a card reader.

2. Conceptual issues

The importance of a clear, unanimous and logically structured operational concept has been stressed above. The tests showed that a consistent naming of all functions is especially important for this specific user group. There were problems with ambiguous terms. Some users were not able to clearly understand and allocate the terms "programme", "channel", "timer" or simply "new" for "new timer entry". They also mixed up "Main menu" and "Ammec menu". Any naming should clearly and consistently describe what is actually meant. This, of course is not only true for blind users but for any user, especially for those who are not as technologically experienced.

Having watched the users when trying to solve and fulfil the test tasks it became clear how important it is that the menu structure is consistent and uniform across all navigation levels. It is quite obvious that especially blind users need to adapt themselves to one certain navigation principle and do not want to switch between different levels and different navigation principles. This implies that all functions should be accessible via the same basic principle.

Due to the context-sensitive concept of colour buttons in the case of the Ammec approach (their functions are different depending on which menu section is actually being used) users need to have these functionalities read out to them over and over again whenever they change the navigation level. Such specific conceptual issues, if they cannot be avoided, need to be thoroughly explained to the users from the start.

3. Orientation and Feedback

The strongest point of criticism, both spontaneously voiced during the task accomplishment but also when replying to our evaluation questions later on, was that users missed feedback from the system. Auditory feedback (feedback sounds for success or even spoken announcements like "action accomplished") are a much wanted feature. Maybe such feedback could be integrated into a user profile as users might be disturbed by it once they have become more familiar with the system. This should even be complemented by further functions for better orientation and options for context-sensitive help functions, all in order to offer orientation and convenient user friendly navigation through the service.

During the tests, users criticised over and over that the current channel was not announced whenever they switched channels or when they had left the menu. Another misunderstanding as the Ammec does actually offer this function which needs to be activated through a certain button.

Programming and deleting a recording was problematic for quite some users. Here they would have liked questions like "Do you really want to delete your timer-entry for programme xxxx? then please confirm with OK" etc….) or they wanted an announcement like: "You have now set your timer for programme xxx at xxxx". When replaying their recorded programmes users missed a help function on how to deal with the player functions. Here, for example, they suggested introducing a recognisable typical sound symbolising winding the film back or forward.

Once the videotext is accessed, it is being read out straight away. This was criticised by testers. A short break and automated reading out of the first line might be convenient for better orientation. This concerns especially users who are not yet used to speech output systems. Some ask for a clearer structure like, for example, starting with a contents page. This however, would again contradict the principle of inclusion. Also when using the videotext the users complained about missing feedback signals, like for example a confirmation of entering a page number.

9.3.4 Remote Control

Concerning the remote control that was used in the tests "poor workmanship" was criticised, including the dissatisfying pressure point of the directional pad, the fact that not all buttons had functions and the fact that the "rocker button was situated too close to the directional pad". A good pressure point and good focused triggering are very important, especially for blind users. Blind users are not always able to be as precise in positioning the remote control. Therefore, the triggering radius should be quite large. The remote control was deemed too large. However, users found the buttons too small. To find an optimum solution to this problem requires compromise. It was stated that there was not sufficient space in between the buttons and that there were not enough haptic points for orientation (like, for example on the "5" button, and the "OK" button). Users also missed orientation assistance (shapes?) for the colour buttons. Shape is an important issue, rocker buttons are easily identified by blind users and are useful for changing sound or volume and for switching channels.

In order to improve recognition of buttons, users asked for a sound signal or an announcement once a button is being pressed, such a function should be individually configurable. Another option for such functionality would be a sound signal of low battery power.

Using the "Star-button" on the remote control for confirming a selection was criticised, users had preferred and expected to use the OK-button to this purpose, In turn, the grouping of buttons, and the fact that certain functions could be directly activated via the remote control were spontaneously praised.

10. Appendices (further Questionnaires etc.)

The questionnaires for testing the Enhanced Text Service and for testing Clean Audio are contained in Deliverable D3.4

10.1 Questionnaire concerning Reduced Playback Speed

Test duration: maximum 3 x 10 min = 30 minutes provided on a DVD

Time per each particular video contribution: about 10 minutes

Video contributions foreseen with genres like: the talk show, sport, science

Methodology:

Three different footages with demanding content A, B and C will be played for about one minute's duration. The test persons will indicate with crosses in a simple grid how well they understand the content. The test persons are dyslexics and some people with cognitive impairments. It is also assumed that both younger and elderly people will be the target for the laboratory tests and they should be included in the tests.

Examination:

It is to be expected that the best understandability (intelligibility) will be achieved by the target group watching videos with the speed reduced to 90% down to 75% of the original speed. All other participants in the tests are expected to find the original speed (100%) preferable.

Possible questionnaire for the "Reduced Playback Speed" tests:

This test offers the same content with duration of about 1 Minute at 3 different speeds: reduced down to 90%, 75%, and 50%, of the original speed, respectively.

Please place a single cross in each line of the grid!

At the beginning of the test the play-out speed is 100%, this means, the original speed.

Valid for all three genres: sport, the talk show, science

	hard to understand	understandable	Easy to understand	Easy to understand but too slow
Contr. A1 Original				
Contr. A2 90%				
Contr. A3 75%				
Contr. A4 50%				

For interpretation: a straight line is expected which goes from “hard to understand” to “good to understand but too slow”.

10.2 Questionnaire concerning Video Signer by RAI/Brunel

Modified questionnaire concerning Video Signing

Parte Scritta	Video
Il progetto DTV4All ti ringrazia per Il tempo e l'attenzione che stai dedicando a questo questionario. *Perché del questionario:* Il questionario serve per comprendere meglio le esigenze dei sordi in materia di comunicazione. Mediante l'analisi delle risposte si potrà valutare la reazione degli utenti in merito ad un servizio di Accesso Emergente per il linguaggio dei segni. Il servizio tradizionale di interpretazione in L.I.S (Lingua dei Segni Italiana) è legato ad alcune edizioni del telegiornale o di programmi dedicati. Lo schermo viene, in pratica, diviso in due per ospita contemporaneamente il conduttore e l'interprete. Queste edizioni risultano poco fruibili dal pubblico udente perché la finestra con l'interprete occupa una porzione significativa del video. Ne consegue una difficoltà oggettiva di inserire il servizio di interpretazione nelle fasce principali del palinsesto. l servizio che si intende valutare permette all'utente di attivare o disattivare l'interprete. La flessibilità dei canali digitali permette anche di definire la posizione e la dimensione della finestra in cui viene visualizzata la traduzione in LIS. Le tue risposte serviranno ad orientare le scelte per la realizzazione e l'erogazione del servizio.	

Ruolo, valore e personalizzazione dell'interprete

Parte Scritta	Parte Video
1. Quando guardi la televisione, ti piacerebbe poter spostare liberamente sullo schermo l'immagine dell'interprete LIS? – Sarebbe molto utile – Sarebbe abbastanza utile – Sarebbe poco importante – Sarebbe inutile	Sample related to the question
2. Quale soluzione preferisci riguardo alla posizione della finestra con l'interprete ? Rispondi dopo aver guardato gli esempi. – In basso a destra – In basso a sinistra – In altro a destra – In alto a sinistra	Sample related to the question showing the interpreter window positioned in each corner
3. Quando guardi la televisione, ti piacerebbe poter liberamente scegliere la dimensione sullo schermo della finestra con l'interprete ? – Sarebbe molto utile – Sarebbe abbastanza utile – Sarebbe poco importante – Sarebbe inutile	Sample related to the question
4. Utilizza la tastiera per impostare la dimensione preferita della finestra in base allo schermo numero uno (dimensione 14") – Dimensione preferita: 10% – Dimensione preferita: 20% – Dimensione preferita: 30% – Dimensione preferita: 40% – Dimensione preferita: 50% – Dimensione preferita: 60% – Dimensione preferita: 70% – Dimensione preferita: 80% – Dimensione preferita: 90% – Dimensione preferita: 100%	The user can change the window size according to their viewing preference

5. Utilizza la tastiera per impostare la dimensione preferita della finestra in base allo schermo numero due (dimensione 28”) – Dimensione preferita: 10% – Dimensione preferita: 20% – Dimensione preferita: 30% – Dimensione preferita: 40% – Dimensione preferita: 50% – Dimensione preferita: 60% – Dimensione preferita: 70% – Dimensione preferita: 80% – Dimensione preferita: 90% – Dimensione preferita: 100%	The user can change the window size according to their viewing preference
6. Utilizza la tastiera per impostare la dimensione preferita della finestra in base allo schermo numero tre (dimensione 46”) – Dimensione preferita: 10% – Dimensione preferita: 20% – Dimensione preferita: 30% – Dimensione preferita: 40% – Dimensione preferita: 50% – Dimensione preferita: 60% – Dimensione preferita: 70% – Dimensione preferita: 80% – Dimensione preferita: 90% – Dimensione preferita: 100%	The user can change the window size according to their viewing preference

Visualizzazione della traduzione su dispositivo palmare

Parte Scritta	Parte Video
7. Quando guardi un programma televisivo, ti piacerebbe avere la traduzione in LIS visualizzata sul tuo telefono cellulare o su un visualizzatore portatile? – Preferisco vederla sul televisore – Vorrei vederla sul cellulare o su un dispositivo portatile – Vorrei entrambe le soluzioni	The user can see the translation into sign language both on the TV screen and on a smart phone.
8. Quando guardi un film al cinema, ti piacerebbe avere la traduzione in LIS del film visualizzata sul tuo telefono cellulare o su un visualizzatore portatile? – Sarebbe molto utile – Sarebbe abbastanza utile – Sarebbe una funzione poco importante – Sarebbe inutile	The user can see the translation into sign language on a smart phone.
9. Ritieni sia maggiormente faticoso seguire la traduzione sullo schermo del palmare invece che sul televisore ? – E' molto più faticoso – E' leggermente più faticoso – Non riscontro differenze – E' meno faticoso – Età	The user can see the translation into sign language both on the TV screen and on a smart phone.

Informazioni personali

Parte Scritta	Parte Video
10.Sesso: – Femmina – Maschio	
11.Scolarità: – Elementari – Medie – Maturità – Laurea	
12.Tu sei sordo: – dalla nascita – da prima dei 3 anni – da dopo i 3 anni	
13.Qual è il tuo deficit uditivo?: – Sordo profondo – Sordastro	
14.Provieni da una famiglia in cui sono sordi: – Entrambi i genitori – Un genitore – Nessuno dei due genitori	

Saluto: Grazie per aver dedicato il tuo prezioso tempo alla compilazione di questo questionario. Ti rivolgiamo un cordiale saluto.	

10.3 Detailed Results of Demonstrator of Text-To-Speech-Applications in the German Language

This section documents the detailed results of the tests with a Demonstrator for Text to Speech Applications in the German language. Like the other laboratory tests performed by RBB, this test was quite complex as it encompassed tasks (indirect method) in the second step of each individual test session and finally evaluations and direct questions as the third part. Different from the other tests, however, the idea of the specific test was not so much to test a novel service or device developed in DTV4All. Instead, testers' feedback on a device already on the market, the so called Ammec, was used to derive generic recommendations as to text to speech applications for blind people. Therefore, only the more generic conclusions were translated into English and are contained in chapter 9 of this document. The tasks as part of the detailed results listed below were translated into English to enable wider understanding.

10.3.1. Evaluation of tasks

Bereich 1 – TV gucken, EPG nutzen

Thema: Umschalten

1. Could you please go to the main menu?

Objectives	**Success**	**With help**	**Failure**
1.1. User recognises the button "Menu" on remote control as general access to the speaking interface as such.	60	40	0
1.2. User understands that he/she can move up and down inside the main menu	30	10	60
1.3. User navigates in Level 1. He/she understands the correlation of the Speaking Interface with remote control (*OK-button, number buttons, arrow keys*)	70	10	20

Die Zahlen zeigen, dass alle den Button „Menü" finden. Allerdings benötigten immerhin 40 Prozent eine Hilfestellung. Mehr als die Hälfte hat nicht verstanden was beim hoch und runter schalten geschieht. Die Menüstruktur ist in der ersten Ebene ist für die meisten Nutzer ersichtlich, nur zwei scheiterten.

2. Please switch to channel RBB

Objectives	Success	With help	Failure
2.1. User understands main menu and use of arrow keys for moving in main menu	90	10	0
2.2. User navigates in Level 2 (Channel list)	100	0	0
2.3. He/she understands the correlation of the Speaking Interface with remote control (*OK-button, number buttons, arrow keys*)	80	10	10

Die meisten Nutzer konnten ohne Hilfe umschalten und haben sich in der zweiten Ebene zurechtgefunden.

3. Please switch to channel ARD

Objectives	Success	With help	Failure
3.1. Repetition: User understands main menu and use of arrow keys	100	0	0
3.2. Repetition: Navigation in Level 2 (Channel list)	90	10	0
3.3. He/she understands the correlation of the Speaking Interface with remote control (*OK-button, number buttons, arrow keys*)	90	10	0

Bei dieser Wiederholung der vorherigen Aufgabe konnten alle Nutzer ohne Hilfe umschalten. Auffällig ist aber, dass die Navigation in Ebene 2 beim zweiten Mal von einem Nutzer nur mit Hilfe bewältigt wurde. (Beim ersten Mal bewältigten alle Nutzer die Navigation ohne Hilfe) Das lässt vermuten, dass die Navigation nicht von allen Nutzern logisch nachvollzogen wurde, sondern „nur“ intuitiv richtig bedient wurde.

A. User Evaluation	Simple	Manageable	Complicated	Too complicated
How easy was it for you to switch channels?	60	30	10	0

Objectives	Success	With help	Failure
4.1. Repetition Understanding main menu	90	10	0
4.2. Navigation in noch unbekannter Ebene 2.	60	40	0
4.3. Erfassen neuer Art von Informationen (Programm-Daten)	70	20	10

Die anschließende Einschätzung der Nutzer fällt interessanterweise nicht so positiv aus. Zwar haben 60 Prozent das Umschalten mit „einfach“ bewertet, aber immerhin 30 Prozent beurteilten diese Funktion mit machbar und ein Nutzer fand es sogar kompliziert umzuschalten.

Beobachtung und Kommentare der Tester zusammengefasst:

Einige Nutzer beklagten sich über die schlechte Ansteuerung der Fernbedienung. Weitere Probleme hatten ihre Ursache in den nicht eindeutigen Bezeichnungen. So kam es zur Verwechslung von Hauptmenü und Ammec-Menü. Einige Nutzer hatten Probleme, die Bezeichnungen Programm, Kanäle, Sender eindeutig zuzuordnen. Die Verwendung der Sternchen-Taste für die Bestätigung wurde von einigen Nutzern, die lieber die OK-Taste entsprechend nutzen würden, kritisiert.

Vielen Nutzern fehlten Rückmeldungen des Menüs, auch beim Verlassen des Menüs oder beim Umschalten. Fast alle Nutzer wünschten sich eine kurze Ansage darüber, wo man sich nach einer Aktion befindet. Vor allem eine Ansage des aktuellen Kanals wurde immer wieder verlangt. (Beispiel: Menü verlassen, Aktueller Kanal/Sender: Arte…) Die Beobachtungen zeigen, dass die meisten Nutzer Probleme haben, die Menüstruktur auf Anhieb zu verstehen. Diesbezügliche Lösungsansätze könnten sein: Eine detailliertere Hilfestellung im Handbuch oder die Möglichkeit bei Bedarf kontextsensitive Hilfeseiten aufzurufen. Ein aufwändigerer Ansatz wäre die Entwicklung eines kleines Lernprogramms, das bei der ersten Nutzung automatisch aufgerufen werden kann.

Außerdem wird die Haptik der Fernbedienung kritisiert, sie sei zwar schön groß, das nütze allerdings nichts, wenn die Knöpfe so klein und schlecht voneinander trennbar seien.

Fazit zum Thema Umschalten:

a) Es werden mehr Erklärungen zum Verständnis der Menü-Struktur benötigt.
b) Allgemein wird das Umschalten als einfach bewertet, zumindest mit ein bisschen Erfahrung.
c) Offensichtliche Schwierigkeiten liegen uneindeutigen Bezeichnungen
d) Eine Ansage des Senders bei Senderwechsel wäre wichtig
e) „Ortsangaben“ wären sinnvoll: Ist man im Menü, in einem Untermenü oder wieder im TV-Modus?

Thema: Zusatzinformationen (EPG)

4. Please access additional information on the RBB-programme that will be broadcast today at 20:15 pm.

Bis auf einen Nutzer haben sich alle im – schon bekannten - Hauptmenü zurechtgefunden. Fast die Hälfte der Nutzer brauchte jedoch Hilfe bei der Navigation in der noch unbekannten Ebene. Rund 70 Prozent der Nutzer konnten die neuen Informationen erfassen und einordnen, 20 Prozent brauchten Hilfe und 10% scheiterten an dieser Stelle.

Die Hälfte der Nutzer bewertete diese Aufgabe als „machbar", 30 Prozent sogar mit „einfach" und jeweils 10 Prozent mit „kompliziert" und „zu kompliziert". Diese Bewertung zeigt, dass die Navigation an dieser Stelle vereinfacht werden sollte und für viele Nutzer der Begriff Programm für diese Funktion nicht eindeutig.

Beobachtung und Kommentare der Tester zusammengefasst:

Die Nutzer kommentierten während der EPG-Nutzung ähnliche Kritikpunkte wie beim Umschalten. Die Fernbedienung wurde erneut kritisiert, an dieser Stelle fiel auf, dass Abbruch und Bestätigung weiter auseinander liegen sollten und dass die Anzahl der Knöpfe übertrieben und zu komplex ist. Hauptkritikpunkt liegt auch hier in der uneindeutigen Bezeichnung „Programm" für Zusatzinformationen. Die Tester schlugen hier die Begriffe „ Programminformation" und „EPG" vor. Zur zusätzlichen Verwirrung tragen wiederum die fehlenden Rückmeldungen bei, fast alle Nutzer waren sich nicht immer darüber im Klaren wo sie sich gerade befanden. Beim Verlassen des Menüs und beim Wechseln einer EPG-Seite wird ein akustisches Signal gefordert. Darüber hinaus wünschten sich einige Nutzer die Option die Uhrzeiten eingeben zu können und eine eigene Farbtaste für das EPG. Die Möglichkeit größere Zeitsprünge zu nutzen wurde positiv erwähnt.

Fazit zum Thema EPG:

a) Ansage des Kanals wäre wichtig, um nicht jedes Mal wenn man in den EPG möchte, kontrollieren zu müssen ob man gerade im richtigen Kanal ist, eine eindeutige Taste dafür wäre hilfreich

b) Ansagen des Senders beim Umschalten auf jeden Fall nötig

c) Eindeutigere Bezeichnungen (insbesondere bei den Begriffen „Programm, Kanäle, Sender") wären gut

d) Farbtaste (Grün) ist im normalen Fernseh-Betrieb mit Programm belegt, das erfährt der Nutzer aber nicht durch die Farbtasten-Vorlese-Funktion, sondern nur durch Zufall

Bereich 2 – Videotext

1. Please start RBB's teletext. Please move to page 112. Then go to the first news item on page 118 and have the news item read out to you.

Objectives		Success	With help	Failure
1.1. Repetition Understanding main menu without any hints		100	0	0
1.2. Navigation in Level 2 (Teletext)		100	0	0
1.3. Understanding the teletext information		100	0	0
1.4. Navigating teletext		100	0	0
C. User Evaluation	**Simple**	**Manageable**	**Complicated**	**Too complicated**
How easy was it for you to use teletext?	100	0	0	0

Der Teletext wurde von allen Nutzern ohne Hilfe genutzt, Inhalt und die Struktur wurden verstanden und auch die Navigationslogik wurde ohne Hilfe angewendet. Dementsprechend bewerteten die Nutzer die Teletext-Anwendung alle mit „einfach".

Beobachtung und Kommentare der Tester zusammengefasst:

Wenn der Videotext aufgerufen wird, wird sofort vorgelesen. Ein wenig Zeit und kurze Infos wo man sich befindet, wären hilfreich um sich zu orientieren. Eine Hervorhebung der Taste 5 auf der Fernbedienung würde die Zuordnung der Zahlen erleichtern. Die „Stumm-Taste" sollte für alle Anwendungen mit Sprachausgabe freistehend auf der Fernbedienung angeordnet sein, das wäre laut einiger Tester sinnvoll. Diese Taste wird zum Beispiel benötigt falls außerhalb des Fernsehers etwas passiert oder der Nutzer den TV-Ton hören und danach im Videotext weiterlesen möchte. Außerdem sollte in diesem Fall ein Fehlerschutz eingebaut sein, z.B. ein Erinnerungssignal wenn im Stumm-Modus weitere Tasten gedrückt werden. (Ähnlich des Geräuschs, bei der Passworteingabe wenn die Feststelltaste beim PC aktiviert ist) Auch während der Videotext-Nutzung wird bemängelt, dass es zu wenige Rückmeldungen gibt - es fehle eine Rückmeldung bei der Zahleneingabe.

Fazit zum Thema Videotext:

a) Videotext wird immer direkt vorgelesen, ohne Aktivierung, das stört die Tester, sie suchen nach einer Stumm-Taste.

b) Trotz des guten Ergebnisses zeigten unsere Beobachtungen, dass die Struktur des Videotextes nicht allen Nutzern klar wurde. Empfehlenswert wäre hier das im Videotext existierende Inhaltsverzeichnis generell als Startseite anzubieten..

c) Die Nutzer wissen nicht ob ihre Zahleneingabe von der Fernbedienung übernommen wurde, hier sollte die jeweilige Eingabe angesagt werden.

Bereich 3 – Aufzeichnungen

Thema: Aufzeichnung/Timer programmieren

1. Please set the Timer for a recording: Channel ARD, Time 20:00, News

Objectives	Success	With help	Failure
1.1. Repetition Understanding main menu	90	10	0
1.2. Navigation in Level 2 (Timer).	60	40	0
1.3. Applying the colour button on remote control.	40	50	10

Die obigen Zahlen im Bereich "Aufzeichnung" zeigen, dass das Hauptmenü zwar von den meisten Nutzern (90 Prozent) verstanden wurde. Allerdings brauchte fast die Hälfte der Nutzer (40 Prozent) Hilfe bei der Navigation in der Ebene 2. Und die Einstellung des Timers schafften nur 40 Prozent der Nutzer ohne Hilfe, 50 Prozent bewältigten die Aufgabe mit Hilfe und 10 Prozent scheiterten.

Beobachtung und Kommentare der Tester zusammengefasst:

Einige Nutzer sind durch die zwei unterschiedlichen Lösungswege im Manual verwirrt, sie bringen beide Wege durcheinander und versuchen sie zu kombinieren. Der Begriff Timer wird nicht von allen Nutzern verstanden. Mehr als die Hälfte der Nutzer ist sich zwischendurch immer wieder unsicher in welchem Programm (Kanal) sie sich gerade befinden. Fast alle Nutzer kritisieren auch an dieser Stelle die fehlende Rückmeldung des Systems. Es sollte eine Bestätigung geben, wenn man im Timer-Menü ist. Einigen Nutzern ist die Eingabe über die Zahlen nicht klar.

Thema: Programmierte Aufzeichnung löschen

2. Please delete the timer entry

Objectives	Success	With help	Failure
2.1 Repetition Understanding main menu	90	10	0
2.2 Repetition Navigation in Timer-Level	70	30	0
2.3 Applying the colour button on remote control.	80	20	0
2.4 Does the user return autonomously to TV and closes the Menu (EXIT)?	60	10	30

Bei der Aufgabe, die gespeicherte Aufnahme wieder zu löschen, zeigt sich ein ähnliches Bild wie bei der vorherigen Aufgabe. Das Hauptmenü wird von der eindeutigen Mehrheit (90 Prozent) der Nutzer ohne Hilfe verstanden. Bei der Navigation in der Timer-Ebene brauchten dagegen 30 Prozent der Nutzer Hilfe. 80 Prozent der Nutzer verwenden die Farbtasten ohne Hilfestellung und die restlichen 20 Prozent bewältigen die Aufgabe mit Hilfe. Und das obwohl die Farbtasten nicht vorlesbar sind und die Nutzer die Farbtasten allein aus der Erinnerung an das Manual bedienen. Den Menüpunkt mit Exit zu beenden fiel auffällig vielen Nutzern schwer. 30 Prozent der Nutzer

scheiterten und 10 Prozent brauchten Hilfe, nur 60 Prozent haben diesen Schritt ohne Hilfe geschafft.

Beobachtung und Kommentare der Tester zusammengefasst:

Die Bezeichnung „neu“ für neuen Timereintrag verwirrte einige Nutzer, grundsätzlich werden längere eindeutigere Bezeichnungen befürwortet. Die Möglichkeit einen Eintrag zu deaktivieren wurde als überflüssig kritisiert, wobei zu bedenken ist, dass diese Funktion für Urlaube/Abwesenheiten vielleicht sinnvoll wäre. Die Kritik an der fehlenden Rückmeldung betrifft die Orientierung diesmal radikal. Den Nutzern fehlt die Bestätigung beim Löschen, die Zeiten und die Timer-Liste werden nicht vorgelesen. So verstehen nicht alle Nutzer die Systemstruktur und einige scheitern.

D. User Evaluation	**Simple**	**Manageable**	**Complicated**	**Too complicated**
How easy was it for you to programme and delete a recording?	30	30	30	10

Die Beurteilung der Nutzer fiel an dieser Stelle negativer als bisher aus. Jeweils 30 Prozent der Nutzer bewerteten das Löschen eines Timer-Eintrags mit „einfach“ und mit „machbar“ und weitere 30 Prozent sogar mit „kompliziert“. 10 Prozent empfanden die Aufgabe als zu kompliziert.

Beobachtung und Kommentare der Tester zusammengefasst:

Nur ein Nutzer bewertet die Aufzeichnung in den Kommentaren positiv, alle anderen üben deutliche Kritik. Einige äußern, sie müssten sich eben nur an das System gewöhnen und dann würde es besser werden. Aber es ist offensichtlich, dass die Rückmeldungen fehlen und die Menüstruktur sehr schwierig zu verstehen ist. Vor allem Bestätigungen beim Programmieren der Aufnahme und beim Löschen fehlen. Nutzer wünschen sich die Möglichkeit, Einträge über den EPG zu löschen, eine Funktion mit welcher wortweise oder buchstabenweise vorgelesen werden kann und einen haptischen Orientierungspunkt zwischen den Farbtasten. Darüber hinaus wäre die Möglichkeit, ein Ausführlichkeitsprofil - von Anfänger-Modus bis Fortgeschrittene - anzuwählen, eine wünschenswerte Weiterentwicklung. Die # für Aufnahme erinnert an DOS, es könnte auch eine gesprochene eindeutige Information geben.

Fazit zum Thema Aufzeichnung programmieren und Aufzeichnung löschen

a) Feedback fehlt wenn man in Timer-Menü ist. Die User wissen nicht wo sie sich befinden
b) Uhrzeiteingabe müsste angesagt werden, mehr Hilfe an diesem Punkt ist wichtig
c) Der Begriff „Neu“ für neue Aufzeichnung programmieren ist nicht eindeutig
d) Einige Tester dachten sie müssten erst die Sendung im Programm suchen mit OK bestätigen und dann den Timer setzen
e) Mehr Rückmeldung wird an dieser Stelle von allen Testern gewünscht!

Thema: Aufzeichnung aufrufen

3. Please look at one of the recordings. Please choose the one that comes second on the list.

Objectives	Success	With help	Failure
3.1 Repetition Navigation in Timer-Level	100	0	0
3.2 Applying the colour button on remote control.	80	20	0
3.3 Does the user return autonomously to TV and close the Menu (EXIT)?	60	20	20

Die Navigation in der Timer-Ebene fiel allen Nutzern beim zweiten Mal leicht und niemand brauchte Hilfe. Wie bei der vorangegangenen Aufgabenstellung nutzen 80 Prozent der Nutzer die Farbtasten ohne Hilfe (aus der Erinnerung), 20 Prozent benötigten dagegen wieder Hilfe. Den Menüpunkt mit Exit zu beenden, fiel auch beim zweiten Mal auffällig vielen Nutzern schwer. 20 Prozent der Nutzer scheiterten und 20 Prozent brauchten Hilfe, wieder haben nur 60 Prozent diesen Schritt ohne Hilfe geschafft.

E. User Evaluation	Simple	Manageable	Complicated	Too complicated
How easy was it for you to view a recording?	90	0	10	0

Die Beurteilung der Nutzer fiel weitaus positiver als bei der vorangegangen Aufgabe aus und das obwohl die Zahlen bei der Bewältigung bei dieser Aufgabe nicht viel positiver waren. Nur 10 Prozent beurteilten das Anschauen einer Aufzeichnung mit „kompliziert“, die restlichen 90 Prozent beurteilten den Vorgang mit „einfach“.

Beobachtung und Kommentare der Tester zusammengefasst:

Viele Nutzer bemängelten, dass eine Rückmeldung fehle wenn die Aufzeichnung abgespielt wird und dass das System zu langsam sei. Wie das Abspielen gestoppt wird, war einigen Nutzern nicht klar, diese hätten das ganze Gerät abgeschaltet und neu gestartet. Wenn man in das Timer-Menü gehe sei man immer auf dem zuletzt ausgewählten und nicht auf dem ersten, so sei es schwierig einen Überblick über alle Einträge zu bekommen. Die Bezeichnung „Aufzeichnungen“ sei verwirrend. Ein Nutzer sucht die Player-Funktionen über die Farbtasten und ein weiterer Nutzer hat Schwierigkeiten das System zu verstehen, insbesondere die Tatsache, dass alles mit OK bestätigt werden muss.

Fazit zum Thema Aufzeichnung aufrufen

a) Feedback wenn die Aufzeichnung gestartet wird wäre wichtig.

b) Player-Funktionen sollten angesagt werden, Play, Vorwärtsspulen und gegebenenfalls mit Ton unterlegt werden, „schnelles spulen = schnelles Geräusch“

c) Bereich 4 - Radio

Thema: Radiosender aufrufen

1. Please switch to Radio Channel „Spreeradio"

Objectives	Success	With help	Failure
1.1. Repetition Understanding main menu	100	0	0
1.2. Navigation in level 2 (Radio)	100	0	0
1.3. He/she understands the correlation of the Speaking Interface with remote control (*OK-button, number buttons, arrow keys*)	90	10	0

Das Radio zu nutzen war für die meisten Nutzer ohne Hilfe zu bewältigen. Im Hauptmenü fanden sich mittlerweile alle Nutzer ohne Hilfe zurecht und die Navigation in der Radio-Ebene verstanden alle Nutzer ebenfalls ohne Hilfe.

F. User Evaluation	Simple	Manageable	Complicated	Too complicated
How easy was it for you to find the radio channel?	90	10	0	0

90 Prozent fanden es einfach einen Radiokanal aufzurufen und 10 Prozent beurteilten die Aufgabe mit machbar.

Beobachtung und Kommentare der Tester zusammengefasst:

Einige Nutzer hatten Schwierigkeiten damit, dass sie den Sender, auf den sie schalten sollten bestätigen mussten, sie dachten, sie hätten mit der Ansage direkt umgeschaltet. Darüber hinaus wurde der Wunsch nach einer übersichtlichen Struktur mit Ordnern und Unterordnern (selbst editierbar oder scannen) nach Anbietern geäußert.

Thema: Radioaufzeichnung/Timer programmieren

2. Please set the timer for a recording. Please choose 21:00-21:15.

Objectives	Success	With help	Failure
2.1 Repetition Understanding main menu	100	0	0
2.2 Navigation in level 2 (Radio timer)	60	40	0
2.3 Applying the colour button on remote control.	80	20	0

Eine Radiosendung aufzuzeichnen war für einige Nutzer nur mit Hilfe möglich. Die Navigation im Hauptmenü verstanden alle Nutzer weiterhin ohne Hilfe. Aber die Navigation in der Radio-Timer-Ebene war für 40 Prozent der Nutzer nur mit Hilfe möglich, und mit der Nutzung der Farbtasten hatten 20 Prozent Probleme und waren auf Hilfe angewiesen.

Thema: Radioaufzeichnung/Timer löschen

3. Please delete the timer entry

Objectives	Success	With help	Failure
3.1 Repetition Understanding main menu	100	0	0
3.2 Repetition Navigation in Timer-Level	100	0	0
3.3 Anwendung Farbtasten (nicht vorlesbar, also Manual oder Erinnerung)	100	0	0
3.4 Does the user return autonomously to TV and closes the Menu (EXIT)?	70	0	30

Bei der Aufgabe, den Timer-Eintrag zu löschen, fielen die Zahlen positiver aus und alle Nutzer fanden sich wie schon seit einigen Aufgaben im Hauptmenü zurecht. Darüber hinaus nutzten alle Nutzer die Farbtasten ohne Hilfe und navigierten ohne Hilfe in der Timer-Ebene. Das Menü zu verlassen gelang dagegen nur 70 Prozent ohne Hilfe, die anderen 30 Prozent scheiterten.

G. User Evaluation	Simple	Manageable	Complicated	Too complicated
How easy was it for you to programme and delete a radio recording?	70	20	10	0

Eine Radioaufzeichnung zu programmieren und wieder zu löschen fanden 70 Prozent der Nutzer einfach, 20 Prozent immerhin machbar und 10 Prozent beurteilten diese Funktionen mit kompliziert.

Beobachtung und Kommentare der Tester zusammengefasst:

Auch an dieser Stelle fehlten vielen Nutzer Rückmeldungen des Systems und teilweise waren die Nutzer auch hier wieder orientierungslos und mussten sich „durchprobieren“. Außerdem wurde angemerkt, dass es unlogisch sei, die Uhrzeit direkt eingeben zu können, sonst müsse man doch immer alles mit OK bestätigen. Eine Korrekturtaste bei der Zeiteingabe wäre hilfreich. Separate Tasten für Radio und TV zum Hoch- und Runterschalten wären schön. Gewünscht wurde ein eigener Menüpunkt für Radio und entsprechend die Zahl 1 für das erste Radioprogramm.

Fazit zum Thema Radio

a) Bei der Anwahl von Radiosendern verwirrt es die Nutzer, dass sie noch mit OK bestätigen müssen
b) Die Nutzer wünschen sich eine Unterstruktur von Radiosendern nach Anbietern
c) Das Thema Rückmeldungen ist auch hier wichtig – diese fehlen und sind vom Nutzer gewünscht
d) Eine Korrekturtaste bei Zeiteingabe bei Radioaufzeichungen wird gewünscht
e) Separate Tasten für Radio und TV zum Hoch- und Runterschalten wären schön,
f) Nutzer sind verwirrt das sie bei der Zeiteingabe für die Uhr nicht bestätigen müssen mit OK.

Evaluation of questions and comments after the task part

1. Allgemeines Feedback Ammec

a. Was gefällt Ihnen am Ammec am besten?

	Anzahl in %
Sprachausgabe an sich	60
Viele Funktionalitäten	30
Aufzeichnen	30
Videotext lesen	20
Zugang zu EPG Daten	20
Verschiedene TV Systeme	10

Klar ersichtlich ist, dass sich die Mehrheit der Nutzer spontan und ohne Vorschläge ein Gerät mit Sprachausgabe und einem großen Funktionsumfang wünscht. Positiv erwähnt werden darüber hinaus die speziellen Möglichkeiten Sendungen aufzuzeichnen und sich den Videotext sowie den elektronischen Programm Guide vorlesen zu lassen.

b. Was gefällt Ihnen am Ammec am wenigsten?

	Anzahl in %
Kein Feedback / Bestätigung	80
Größe (Ammec bzw. Fernbedienung)	40
Stimme zu unnatürlich / Klangfarbe zu dumpf	20
* Als Bestätigung ungewöhnlich	10
Bedienerführung	10
Benennung der Menüeinträge	10
Fehlende Möglichkeit Daten abzugreifen	10
Kein Card Reader	10
Keine Taste zum stumm schalten	10
Mangelnde Hilfe	10
Preis	10
Zu langsam	10
Zu viele Funktionen	10

Der größte Kritikpunkt ist mit Abstand, dass das System aus Sicht der Nutzer nicht genug Rückmeldungen gibt. Die Stimme der Sprachausgabe ist für 20 Prozent der Nutzer verbesserungswürdig. Weitere spontan - freie Kritikpunkte sind die nicht - eindeutigen Benennung der Menüeinträge, die mangelnde Hilfe und sogar „zu viele Funktionen“. Diese Defizite (aus Sicht der Nutzer) waren ja bereits auch bei der Ausführung der Testaufgaben sehr deutlich geworden. Ausreichende Bestätigungen, Orientierungshilfen und kontextsensitive Hilfestellungen sind bei der Entwicklung einer Anwendung mit Sprachausgabe unerlässlich. Nur so kann sicher gestellt werden, dass sich blinde Nutzer in dem Menü orientieren können und die Funktionen des Systems richtig nutzen können.

Sowohl der Ammec als auch die Fernbedienung sind den Nutzern zu groß. Das Sternchen-Symbol für die Bestätigung und das Fehlen einer „Stumm-Taste“ wurden ebenfalls negativ bewertet. Bei der Entwicklung eines solchen Systems sollte daher auf die speziellen Bedürfnisse der Zielgruppe auch im Hinblick auf die Fernbedienung beachtet werden. Bei dem Gerät selbst wurde der Preis, die langsame Reaktionszeit und dass Fehlen eines Card Readers bemängelt. Alles Hinweise, die generell bedenkenswert sind wenn es um die Einrichtung einer Sprachausgabe

Außerdem wurde die Synthetische Stimme kritisiert. Daraus lässt sich schließen, dass es förderlich wäre ein solches System um eine Auswahlmöglichkeit zwischen mehreren Stimmen zu erweitern.

Allerdings muss bemerkt werden, dass nur solche Tester die Stimme kritisierten, die erst später im Leben erblindet waren und nicht an die Nutzung einer Sprachausgabe gewöhnt sind.

2. Allgemeine Einschätzung Sprachausgabe

	ja	nein
Würden Sie eine Sprachausgabe für den Fernseher nutzen?	100%	0%

Die allgemeine Einschätzung zeigt ein ganz klares Bild. Die Sprachausgabe für den Fernseher würde von allen Nutzern verwendet werden. Das Bedürfnis der Zielgruppe nach einem solchen System wird durch diese Eindeutigkeit untermauert.

Einstellungen Sprachausgabe	sehr wichtig	unwichtig	wichtig
a) Wie wichtig finden Sie, dass der Nutzer die Lautstärke der Sprachausgabe einstellen kann?	100%	0%	0%
b) Wie wichtig finden Sie es, dass die Lautstärke der Sprachausgabe unabhängig von der Lautstärke der Audio/Video-Wiedergabe einstellbar ist?	100%	0%	0%
c) Wie wichtig finden Sie, dass der Nutzer die Geschwindigkeit der Sprachausgabe einstellen kann?	90%	0%	10%

Die Lautstärke der Sprachausgabe muss vom Nutzer reguliert werden können. Dies muss unabhängig vom Fernseh-Ton möglich sein, da sind sich alle Nutzer mit der Bewertung „sehr wichtig“ einig. Auch die individuelle Regulierung der Sprachgeschwindigkeit ist fast allen Nutzern „sehr wichtig“, nur ein Nutzer bewertet diese Funktion etwas schwächer mit wichtig. Diese Möglichkeiten bietet der Ammec.

3. Fernbedienung

a) Bewerten Sie die Haptik der Fernbedienung.

	zu klein	genau richtig	zu groß
Größe:	0%	0%	10%
Tastengröße:	80%	10%	10%

Auch die Bewertung der Fernbedienung zeigt ein eindeutiges Urteil: die Gesamtgröße der Fernbedienung wird von allen Nutzern mit „zu groß“ bewertet. Die Tastengröße dagegen wird von einer deutlichen Mehrheit mit „zu klein“ bewertet. Dieses Ergebnis unterstreicht die speziellen Anforderungen der Zielgruppe bei der Konzipierung einer nutzerfreundlichen Fernbedienung. Diese werden in den folgenden Kommentierungen genauer spezifiziert.

b) Kommentieren Sie Ihre Erfahrungen mit der Fernbedienung des Ammec

	Anzahl der Nennungen in %
Nicht belegte Tasten sind überflüssig	10
Gruppierung der Tasten gut	30
Schlechte Verarbeitung	20
Abstände zwischen den Tasten könnte größer sein	10
Wippen zu nah am Steuerkreuz: Verwechslungsgefahr	10
Mehr Orientierungspunkte (auf 5, OK Taste)	30
Zu groß	10
Direkte Sprachausgabe auf Fernbedienung wäre gut	10
Nummern zu klein bzw. zu nah	10
Steuerkreuz schwer zu bedienen, Druckpunkt schwierig	10
Funktionen direkt anwählbar mit Taste	10
Tasten gut abgesetzt	10

Die Gruppierung der Tasten wurde von 30 Prozent der Nutzer positiv bewertet.

Bemängelt wurde vor allem dass es allgemein zu wenig Orientierungspunkte (zum Beispiel auf der 5 und der OK-Taste) gibt. Auch wünschen sich die Nutzer einen größeren Abstand zwischen den einzelnen Tasten. Darüber hinaus wurde die schlechte Verarbeitung bemängelt. Dazu gehört auch die Kritik am schlechten Druckpunkt des Steuerkreuzes, nichtbelegten Tasten und dass die Wippen „zu nah am Steuerkreuz“ liegen. Überdies wurde der Wunsch nach einer direkten Sprachausgabe an der Fernbedienung geäußert.

c) Was ist Ihnen an einer Fernbedienung wichtig?

	sehr wichtig	wichtig	unwichtig
Tastbare Symbole auf den Tasten	10%	10%	80%
Tonsignal bei Tastendruck	20%	60%	20%
Vibration bei Tastendruck	10%	10%	80%
Sprachausgabe bei Tastendruck	20%	30%	50%

Den Nutzern wurden verschiedene Möglichkeiten vorgeschlagen, um die Tasten einfacher zu erkennen. Am positivsten wurde ein Tonsignal bei Tastendruck bewertet. Tastbare Symbole und Vibrationen wurden überwiegend mit unwichtig bewertet.

Welche Funktionen an einer Fernbedienung sind für Sie wichtig?

	sehr wichtig	wichtig	unwichtig
nächster und vorheriger Kanal	70%	30%	0%
lauter und leiser	100%	0%	0%
stumm	90%	10%	0%
ein/aus	40%	10%	50%
Zahlen	100%	0%	0%
Videotext	40%	40%	20%

Zu den Standard-Funktionen, die direkt mit der Fernbedienung aufgerufen werden können, sollten lauter und leiser, stumm und die Zahlen gehören. Den nächsten und vorherigen Kanal anwählen zu können wurde ebenfalls überwiegend mit „sehr wichtig“ bewertet. Der Videotext wurde eher mit „wichtig“ als mit „sehr wichtig“ bewertet und die „ein/aus“-Funktion wurde sogar eher mit unwichtig bewertet.

4. Welche Funktionalitäten sollte Ihre Sprachausgabe für den Fernseher bereitstellen?

	sehr wichtig	wichtig	unwichtig
Programminformationen	80%	20%	0%
Videotext	50%	40%	10%
Sendungen aufzeichnen	60%	40%	0%
Aufzeichnungen schneiden	60%	30%	10%
Radiosender wechseln	70%	20%	10%
Radiosendungen aufzeichnen	50%	40%	10%
CD abspielen	10%	60%	30%
DVD abspielen	50%	50%	0%

Fast alle zur Auswahl gestellten Funktionen wurden mit sehr wichtig bewertet. Am wichtigsten ist den Testern die Möglichkeit, Programminformationen aufrufen zu können, dicht gefolgt von dem Wunsch Radiosender wechseln zu können. Sendungen aufzeichnen und schneiden zu können wird immerhin von 60 Prozent der Tester als sehr wichtig bewertet. Und auch der Videotext und die Möglichkeit Radiosendungen aufzeichnen zu können, bewerteten immerhin die Hälfte der Tester mit sehr wichtig und nur jeweils 10 Prozent mit unwichtig. Das Abspielen von CDs und DVDs wurde eher als wichtig als als sehr wichtig beurteilt, bei allen anderen Funktionalitäten war der Wert für sehr wichtig größer als der für wichtig.

10.4 Matrix: User Group for testing Enhanced Text Service at RBB

		Tester 1	T 2	T 3	T 4	T 5	T 6	T 7	T 8	T 9	Sum
Eye Disease	**Remark**										
Diabetic Retinopathia	Retina problems due to Diabetes					x					1
Glaucoma									x		1
Cataract					x				x		2
Retinitis Pigmentosa	Degeneration of the retina		x				x				2
Nystagmus	Trembling of the eye				x		x				2
Age-related macular degeneration	Damage of the retina in older age	x								x	2
Achromatopsia	Total Colour Blindness			x							1
Partial Colour Blindness	Difficulties especially with shades of red/green or blue/yellow						x				1
Myopie	Short-sightedness							x	x		2
Albinism	Low vision and sensitivity to light						x				1
Visual Impairment Level according to WHO											
Level 1		x	x					x			3
Level 2				x	x	x	x				4
Level 3									x	x	2
Visual Impairment as a result of eye diseases											
Visual field loss combined with low vision, tunnelvision			x					x	x	x	4
Night blindness			x								1
Limited contrast sensitivity			x			x					2
Stronger sensitivity to light				x			x		x		3
No stereoscopic vision								x	x		2
Inability to see colour, combined with limited vision				x							1
Blurred vision					x				x		2
Full colour blindness						x					
Problems with adjustment between bright and dark			x				x				2
Partial Colour Blindness							x				1