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.:	**D3.6**
Deliverable title:	**Final Report on Expert User Tests of Emerging Access Services**
Nature of deliverable:	**Report**
Dissemination level	**Public**

Due date:	**31.10.2010**
Actual delivery date:	**15.11.2010**

Document title and history:			
D3.6 – Final Report on Expert User Tests of Emerging Access Services			
Version no:	Date:	Changes	Contributor
001	28-07-2010	First draft distributed for discussion	IRT
002	31-08-2010	Contributions by UAB/TVC	UAB/TVC
003	5-11-2010	Contributions by RBB	RBB
004	10-11-2010	Final compilation by IRT	IRT
005	12-11-2010	Translation of 40 pages in Catalan into English and minor editorial revisions	Brunel
006	13-11-2010	Revision of Catalan translation	UAB
007	15-11-2010	Final editorial check and minor revisions	Brunel

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

Contents

1. Executive Summary

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. Tests undertaken by TVC and supported by UAB are described here in this document. The services' usefulness received user approval, and audio description were very highly rated, not only as a tool for making content more accessible but also as a way of providing solutions for families with different needs. Given the above data, we can conclude that the proposed services have been given a high level of approval by consumers, that the services are easily accessible, and that their quality is high enough so as to consider launching them.

For most of the prelingual deaf, sign language is the essential means to get information. That's why the need for sign language on television should not be underestimated, even though the targeted audience is comparatively small. An individually selectable sign language service broadcast via the IP-channel seems therefore as an attractive option which also saves bitrate and hence costs for on-air delivery. Such a service can be accessed via hybrid set top boxes which have an additional Internet connection. The new hybrid receivers currently emerging on the market are not able to decode two parallel video streams which is an important pre-requisite for a truly hybrid broadcast-broadband solution where only the signing video is broadcast via the IP channel and then combined with the broadcast video on the actual device. Therefore IRT and rbb devised an HbbTV-based-solution where the user can access a video with a built-in signing video on demand via the IP-connection of a hybrid device. rbb's laboratory tests documented in this deliverable focused on two objectives: At first the users were to test the usability of this HbbTV-based demonstrator. The second part was focused on user feedback on important parameters for sign language interpreters on TV, mainly in preparing future truly hybrid solutions. The parameters were synchronicity, image size, video bitrate and video positioning. Summed up, the novel HbbTV-video on demand solution was welcomed very much by the testers and usability was satisfying. Users demand individual settings for adapting the video image size and position. The majority of the testers support a complete simultaneous synchronicity between signer-Video and main broadcast video where even the "real world delay" is not there anymore.

2. Introduction

This deliverable covers the laboratory tests of the remaining two project demonstrators. One of them deals with enhanced Audio Description services. Test preparations were described in D3.5. The tests and test results are included in this document in section 3.

The second part of this deliverable covers rbb's user tests on a signing solution for so called hybrid devices, the "phase II demonstrator". A previous solution (phase I) based on the project "SAVANT" was described in D3.5. The phase I demonstrator showed how a signer video was transmitted synchronously with the broadcast of the programme being signed via a separate broadband Internet connection. A sample of this prototype was transferred to RAI which undertook laboratory tests and user tests already described in deliverable D3.5. Because of the unsatisfying situation where both broadcaster and receiver sides need specially designed software under IPR constraints and the high costs of implementing that solution at the broadcaster another unsynchronized solution was explored in phase II.

The phase 2 demonstration which was implemented through the cooperation of IRT and rbb is an HbbTV-based solution. Originally the partners aimed at implementing a "truly hybrid" solution where the signer video is transmitted synchronously with the broadcast programme over the Internet via a separate Internet link. This plan was modified because the new hybrid receivers are not yet able to decode two parallel video streams and to synchronize multiple video streams even if they are delivered on the same network. Therefore IRT and rbb's solution demonstrates how the user can access a video with a *built-in signing video* on demand via the IP-connection of a hybrid device, "an all-IP solution". The deliverable documents the results of testing this application with ten deaf users. In the second part of its user test rbb tested important parameters for signing solutions on TV in order to obtain results especially for future hybrid solutions.

3. Results of tests with a Demonstrator of Enhanced Audio Description

Partners involved: TVC, UAB

3.1 Introduction:

Audio Description (AD) is an additional audio track with narration for all, but especially 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, this is 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.

3.2 Mature versus emerging AD service

As explained in deliverable D3.5, TV Catalonia has established a regular service containing AD with the help of DVB-T transmitters in the whole area of Catalonia, as well as a feedback service with users both at association level and with individuals –who write regularly offering their opinions of the service and much valued feedback. Tests on "mature services" were undertaken by TVC and UAB and the results were published in D2.4.

It was expected that the results of the user interviews which were undertaken by TVC and UAB, and were reported in D2.4, concerning AD mature services would also be 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 independent of the way it is delivered.

Several tests were conducted in order to validate these existing assumptions. Again, the tests were focused on evaluating usability, usefulness and quality of the audio description services. In this deliverable we present the results and conclusions of the evaluation tests. As each of the different scenarios implies a differentiated service, the results are presented individually for each of the 5 evaluated scenarios. However, prior to presenting the results, we will review the 5 emerging AD scenarios that were evaluated, and then we will sketch the methodology of the procedure for collecting the data.

3.3   The 5 scenarios of enhanced audio description service

TVC sketched five possible future scenarios for reception of an enhanced audio description service. The respective prototypes were developed in order to prove their technical feasibility and, more importantly, in order to be able to conduct a battery of tests over these prototypes. (Section 3.4*)*

An extended report on this work can be found as an annex to the present document.

Previous deliverable D3.5 explains in great detail the nature and purpose of these five emerging scenarios. Here after, in this section, we will only summarise their main features:

1. Live streaming Internet TV: This scenario emulates the live streaming IP broadcasting of digital TV media content with audio description. The contents were broadcasted from TVC's servers to the UAB laboratory.
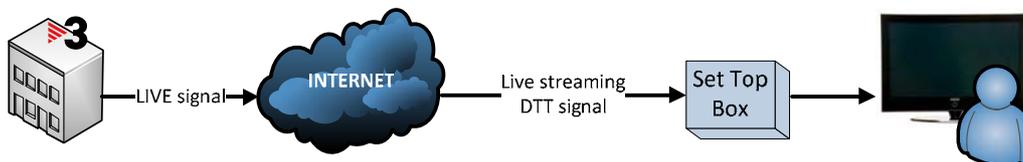

Figure 1 Live Streaming Internet TV distribution

2. AD individual versus collective reception: This scenario combines the simultaneous reception of standard media content with the reception of AD enriched content in the same room. It aims at proposing a solution for a group or family (when watching TV) in those households with a member with visual impairment. Although it is strictly not an emerging service (it is currently available), it has been considered as an emerging service because it is an unusual scenario. For the tests, participants were given a laptop with a Digital Terrestrial Television (DTT) USB tuner that was used to listen to the AD enriched audio thanks to headphones. The image was directly seen on the TV screen.
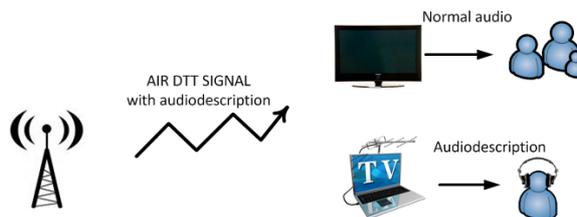

Figure 2 Simultaneous AD / non AD content reception scenario

3. <u>On-demand AD (received on a set-top box)</u>: This scenario lets the user select AD enriched contents on demand amongst a list of programs. The user can then receive them on a TV screen thanks to the set top box, as the data is streamed via the Internet.
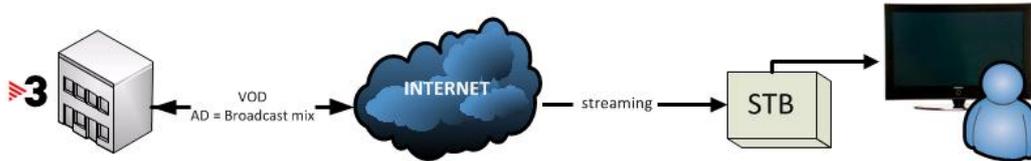
Figure 3 Video on Demand with AD (broadcaster mix) distribution

4. <u>On-demand AD (received on a PC)</u>: In this scenario, participants were asked to download content files on a PC or a laptop in order to view them on the same computer.
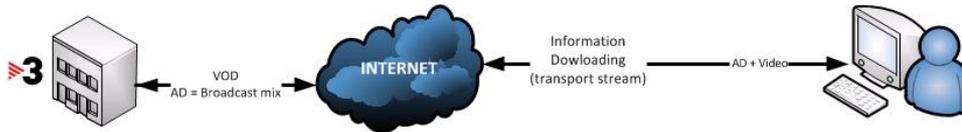
Figure 4 Distribution of AD enriched contents by downloading

5. <u>On-demand AD (podcast reception)</u>: This scenario lets the user download audio files (which are enriched with audio description) on a PC. This way, the user can listen to them via the same computer, or in any other portable device. A laptop with VLC player software was used for the reproduction of the test contents.
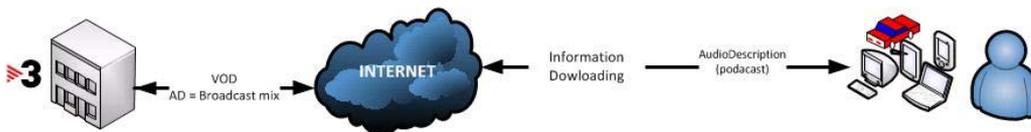
Figure 5 Podcast distribution

3.4   Methodology of the tests

The selection of participants followed the recommendations set by the ITU on methodologies for the subjective assessment of quality in television pictures (ITU-R BT.500-11). Participants were not experts, as they neither work in the area of quality in multimedia systems nor are consultants in this field. The test participation was voluntary. The sample of participants was 14 people (6 women and 8 men). Participant ages were between 21 and 40 years, with an average age of 28.8 years. They understand and speak Catalan and use computers on a daily basis.

Specific criteria were followed when selecting volunteers. Thus, in order to measure vision and hearing, they were subjected to several standard tests. Regarding visual tests, visual acuity was assessed with a computerized version of the Landolt test (Figure 6). Similarly, the possibility of the existence of colour blindness was evaluated with a printed version of the Ishihara plates (Figure 7). Finally, regarding auditory perception, participants answered a questionnaire about everyday situations that can be indicative of potential hearing loss.

Figure 6 Landolt test. A "letter C" of a certain size is shown to participants and they must indicate in which direction the opening is presented. In each trial, the size of the stimulus is reduced. The size at which it ceases to receive the opening indicates the degree of visual acuity.[1]
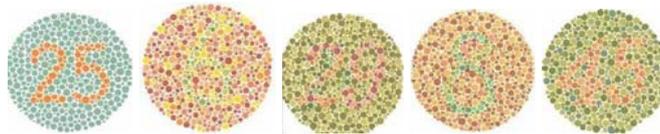
Figure 7 Plates of Ishihara. A plate is presented every time, and the participant must indicate which numbers appear in the figure. Depending on whether you have any colour blindness or not, you will see a figure or another, or none**.[2]**

Additionally, participants were surveyed with the aim of defining a profile of the sample in relation to the habits and preferences when consuming audiovisual content.

The procedure for collecting the data was not easy or simple. The reason for this is that the evaluation of some scenarios had to be undertaken under a time window, as these scenarios were linked to the broadcasting of live content (AD enriched) by TVC. As a consequence, planning for organizing the tests was designed in order to maximize the impact of each session:

---

[1] http://www.precision-vision.com/index.cfm/product/110/logarithmic-landolt-c-eye-chart.cfm [consulted 14/11/2010]

[2] http://colorvisiontesting.com/ [consulted 14/11/2010]

two participants were requested per session, and each participant did the tests in a different order, so that some visualization periods could be shared.

The participants were given information under the scope of a certain protocol. In fact, several documents were written detailing a set of instructions and the purpose of the test, and some specific verbal instructions where established for interacting with the volunteers. Once participants arrived at the laboratory, they were properly informed about the objectives and tasks to be performed through a written document, they also filled in UAB ethical forms. Once they resolved any possible doubts and having verified that the participant had correctly understood the purpose and its work, participants were asked to sign the consent form of participation, as part of the ethical procedure. At this time, the subjective evaluation session began. At all times the volunteers were accompanied by the leaders of the test.

The measuring instruments that were used for the subjective assessment of TVC's emerging services accessibility are a set of five questionnaires, each of which has been adapted to the scenario to be assessed. The purpose of these questionnaires is to gather, in an organized manner, the indicators related to the psychological factors that are involved in the test objectives. That is, questionnaires helped us to obtain information from participants about their attitudes and opinions about a particular aspect that must be assessed but cannot be measured directly. In this sense, the questionnaires consist of three scales, which assess three different aspects of the accessibility emerging services: Usability, Utility and Quality of the received signal.

All details about the sample of participants, the criteria for the selection, the procedure for collecting the data, and the instruments that were used for measuring the tests, can be found in the attached *Annex Subjective evaluation of the enhanced AD emerging services report*.

3.5   General conclusions of the subjective tests on enhanced AD services

The results of the subjective assessment prove that the services tested in this trial are viable. The numerical data collected, together with the users' opinions, indicate what information should be taken into account when planning the deployment of these services.

Table 1 presents a summary of the main features of the services that were rated by the participants as positive or as needing improvement. The green dots indicate positive ratings. The orange dots show the features considered acceptable but with room for improvement. Finally, the red dots indicate features that could be critical when considering the deployment of the service.

Features rated	Scenario 1	Scenario 2	Scenario 3.1	Scenario 3.2	Scenario 3.3
Usability	●	▬	●	●	●
Ease of use	●	▬	●	●	●
Graphic interface	●	▬	●	●	●
Usefulness	●	●	●	●	●
Quality	●	●	●	●	●

Table 1 Summary of the outstanding features in each scenario. Green dots indicate that the variable was evaluated positively in a specific scenario. Orange dots indicate that certain aspects, not decisive for the deployment of the technology, can be improved. Red dots indicate the presence of fundamental aspects that need improvement.

When examining the data in the table, several patterns emerge. Usability, ease of use in particular, was rated highly across the board. It is also interesting to note that the graphic interface was evaluated as needing improvement, especially when including sound descriptions in the menus. The lowest ratings were for the quality of the systems that use streaming. Although it is important to keep in mind that the services evaluated are not comparable, the patterns that emerge from the data are real and the ratings for specific features are consistent (e.g., ease of use). This fact makes the results obtained much more reliable and credible since the participants systematically had similar opinions.

Now a general overview of the features of the different services that were rated by participants and that are summed up in Table 1 is presented.

With regard to the assessment of Scenario 1, Live streaming Internet TV, we can conclude that the system's usability and usefulness are good, while content reception quality must be improved, since the picture is sometimes affected by glitches. Despite this problem, having the same live DTT broadcast available on the Internet was seen as highly useful.

In the case of Scenario 2, AD reception, both usefulness and quality were generally satisfactory. Since the participants without vision impairments were not inclined to use content with AD, the personalized TV viewing system for the visually impaired (via a portable PC) was rated as highly useful. This personalized option, which could suit individual preferences within the family, was rated as the most positive feature by the participants. It was suggested that a smaller device for DTT reception with audio description be used for user convenience.

On scenario 3.1, On-demand AD via a set-top box, usability and usefulness were highly rated. However, aspects of the quality of content received via streaming must be taken into account to try and eliminate picture break-ups and make reception quality as good as the DTT signal. Nevertheless, the participants also pointed out the usefulness of being able to access previously broadcast content at any time.

The evaluation of Scenario 3.2, On-demand AD via a PC, was highly favourable both from the viewpoint of usability and usefulness. Being able to access content at any time and from any place was most appreciated. The rating of quality was affected by content download time (dependent on Internet access bandwidth). Some participants suggested adapting the content for use on mobile devices.

Finally, for Scenario 3.3, On-demand AD via podcast, the data indicates that usefulness was rated a little lower than in the previous scenario mainly because the participants found it hard to follow the series with the audio alone (podcast), even with the help of audio description. Apart from that, both usability and quality were rated very high.

In short, the emerging services analyzed in this trial are technically viable, but some aspects of the quality of the service must be improved. The usability of these services, with respect to ease of use, was clearly validated by the participants in the trial and some suggestions were made for improving the graphic interface. The services' usefulness also received user approval, and audio description was very highly rated, not only as a tool for making content more accessible but also as a way of providing solutions for families with different needs. Given the above data, we can conclude that the proposed services have been given a high level of approval by consumers, that the services are easily accessible, and that their quality is high enough so as to consider launching them.

4. Results of acceptance tests of Video Signing

Partners involved: RBB, IRT

4.1 Introduction

Originally rbb and IRT planned to test a broadcast/broadband approach, where the signer video is transmitted synchronously with the transmission of the programme being signed over the Internet via a separate Internet link where the Phase II Client comprises a Hybrid Digital Video Broadcasting – Satellite (DVB-S)/Digital Subscriber Line (DSL) receiver with a Consumer Electronics - HyperText Markup Language (CE-HTML) browser providing simultaneous decoding of the two video streams, i.e., optional synchronisation of the signer and the programme. This plan was modified due to the following reasons: the new hybrid receivers currently emerging on the market are not able to decode two parallel video streams, an important pre-requisite for a hybrid broadcast-broadband solution. Furthermore they do not support synchronization of multiple video streams even if they are delivered on the same network. Therefore the objective and set-up of the test was modified:

The aim was to evaluate the acceptance of a video-on-demand based signer solution and its basic specifications, also for future hybrid solutions, in order to be able to create signer-on-demand-guidelines for the future enhancement of current TV programmes towards a barrier-free access for deaf people.

As in rbb's three other laboratory tests, the users were recruited with the assistance of regional impairment associations in Berlin and Brandenburg.

In more detail, the laboratory test focused on two strands of work:

- Strand one: An HbbTV-based demonstrator was implemented through the cooperation of rbb and IRT. It showed an all IP-solution where the user can access a video with a built-in signing video on demand via the IP-connection of a hybrid device. The users were to test this HbbTV application in order to find out a) how they would generally like such an approach and b) how user friendly and suitable the application as such is. For preparing the demonstration, rbb engaged a signer and produced an ARD Tagesschau broadcast (main German news programme, 15 minute duration) with a built in signing translation, encoded as a single video file. The graphic design of this was modelled on the current signing solution broadcast daily on channel Phoenix. With this, IRT prepared a DVB transport stream and built in signalling of a specifically developed HbbTV application, which offers the possibility to access the video file via a web server. This transport stream was delivered

by a stand-alone DVB-S playout server to the HbbTV receiver (HUMAX iCord HD+) that was used for testing.

When the Tagesschau news programme started at the receiver, the application was called automatically and simply showed a text popup message, offering the user the possibility ("push the RED button") to watching this programme on-demand over IP with the incorporated sign language interpreter.

Figure 8 Popup text of the HbbTV application

Figure 9 All-IP signer on demand

- Strand two: In order to prepare a "truly hybrid" solution where the main broadcast video and the signing video are transmitted via separate channels but also in order to find out generally about users' needs and ideas about the design of signer solutions for digital TV, this strand focused on user feedback on important parameters. The parameters were synchronicity, image size, video bitrate and video positioning. Here IRT prepared a number of video samples by encoding the main broadcast video combined with the prepared signer video for testing different variants of synchronicity, image size and video bitrate. For checking the signing video position several screenshots were created. All these samples were then processed into a DVD for running the tests.

The tests were carried out at rbb's Berlin premises on 5th and 6th of July 2010 with 10 deaf test persons. Each individual test session took about 1 to 1.5 hours.

4.2   Methodology

4.2.1 Method for Strand One

The method for evaluating the on demand HbbTV-application was both indirect and direct. Indirect: The user had to fulfil a task referring to handling and usability of the signer solution. The meaning and objective of this task was clearly defined. The users' success or failure in performing the task provided some evidence if the exemplary technological approach of the signer service is convenient and user friendly. Furthermore, free spontaneous comments of the users showed which aspects they liked or disliked. Direct method: After fulfilling the task the users were asked some questions to evaluate their test experience and to find out about wanted qualities and functionalities.

4.2.2 Method for Strand Two

For the test, four aspects were selected as representing the most important basic characteristics or parameters for a hybrid solution:

Synchronicity:

The test wanted to explore the best solution in terms of user acceptance for synchronising the main and the signer videos. The basic question was, how exactly do the main programme and the signing video need to be synchronised? Due to video encoding and further factors the DVB-S/-T signal (the main broadcast programme) will lag some seconds behind the IP-distribution while the IP-distribution (the signing video) will undergo slight fluctuations. Consequently synchronicity of the two different video streams will not be guaranteed. Consumer Electronics manufacturers must be able to implement video buffers in such access service enabling devices

and calculate the buffer size for an appropriate synchronisation approach. This test was to be the basis for a future calculation of buffering. For the test, a DVD providing designated example video clips was prepared.

Three test scenarios on three different variants of synchronicity were tackled:

1 *Signing video is delayed compared to main programme.* Starting point is the usual human factor-based delay (mental processing and executing gestures) of sign language output regarding the main input and their video presentation. The question here was: Up to which point is the delay acceptable for users? In order to find out about this acceptance threshold a video sample was prepared, starting a synchronous ("real world"-delay see above) delivery of main and sign language video, and then introducing an increasing delay of the sign language video. This was achieved by using a slightly lower frame rate of 24 fps for the signer and 26 fps for the main video. The test person had to give a sign at the point when the sign language and the main video became incoherent. This point in time was noted.

2 *Signing video is exactly simultaneous to the main broadcast video*, i.e., even the real world delay is not there anymore. The approach is to technically synchronise both videos and remove all delay, so that the sign language sentences fit to the main video speech. Here, both videos were synchronised manually and re-encoded, so the test person had to judge if this approach makes understanding easier. The sign language video was moved in relation to the news clip, i.e. in the result the signer starts when the news speaker says the first word of a new news item.

3 *Signing video is ahead of the main broadcast:* Subtitle tests in DTV4All showed that users even liked subtitles being *ahead* of the main broadcast video, therefore this option also was to be tackled for sign language. Here, the question is, will the acceptance increase if the user is able to access the sign language earlier than the main video? Therefore, the third variant was to technically delay the main video, so that the sign language video is ahead. In the future this might be the case e.g. if a sign language video will be streamed via IP and added to the broadcast main video at the end device, the IP stream will be several seconds ahead of the main video due to the encoding and multiplexing induced broadcast delay. To find out the acceptance threshold a video sample was prepared, starting a synchronous play of the main and sign language video, and then introducing an increasing delay in the sign language video. This was achieved by using a slightly higher frame rate of 26 fps for the sign language video than the 24 fps for the main video. Again, the test person had to give a sign at the point when the

sign language and the main video became incoherent to them, or when they became irritated by the delay. This point in time was noted.

Video bitrate:

In order to evaluate the required image quality and the corresponding video bitrate for a future hybrid signer solution and a given image size, the use of a most current state-of the-art video encoding is necessary. This was tested with the help of a DVD providing designated example video clips. These video samples (having dissimilar language content) were prepared, encoded with H.264 at a fixed resolution of 277x272 pixels and a video bitrate of

a) 200 kbit/s

b) 300 kbit/s

c) 450 kbit/s

300 kbit/s was agreed as the standard medium bitrate for the given resolution. Test persons had to rate the signer's image quality with school marks from 1 (very good) to 5 (very bad).

Image size:

Another objective was to find out the image size for the sign language video preferred by users. Again, different video samples were prepared, encoded with H.264 at a fixed video bitrate of 300kbit/s and an image size of

a) 208x204 pixels

b) 277x272 pixels

c) 416x402 pixels.

277x272 pixels was agreed upon as the standard image size, based on the current signer enhanced news show "Tagesschau" at channel "Phoenix". Test persons had to rate with school marks from 1 (very good) to 5 (very bad) while watching the sample clips off a test DVD.

Positioning:

The aspect of the positioning of the sign language interpreter on the screen is important for raising the acceptance of future hybrid sign language solutions. Here one can expect basically two different approaches for a variety of positions on the screen

a) sign language video overlaps main video

b) sign language video adjoins main video

The test was realised by providing the test persons with screenshots. Again, they had to judge with school marks from 1 (very good) to 5 (very bad) for both overlapping and adjoining the following variations:

- left top
- left middle
- left bottom
- right top
- right middle
- right bottom

4.2.3 Procedure

The users and sign language interpreters were welcomed and introductions made. They were offered refreshments and it was explained to them, in simple terms, what was to happen, that we were testing an exemplary technology and that our ultimate aim is to find out if such a similar technology is of use to them and what it should look like according to their opinions. (This took 10 minutes).

To help put the user at ease, the interviewer chatted with him or her for a couple of minutes. The interviewer explained the test procedure. (This took 5 minutes).

The user was then asked some pre-task questions about computer, TV and corresponding sign language usage. This was to find out about media and computer literacy. The testers were encouraged to say whatever they thought that would help if they said aloud what they thought, the more feedback the better. They were also encouraged to say what they do not like because that would also be very helpful. (This took 10 minutes)

The users were then introduced to the idea of using an interactive HbbTV service and asked to fulfil the task. The interviewer observed and guided the tester through the task, giving help only if necessary. The task was explained in advance and the users were asked if they had understood it. The interviewer(s) observed the tester, noting down any comments made and questions asked. The users were gently encouraged to voice their opinions. (This took 10 minutes).

After that the parameters concerning the four aspects deemed valid for future hybrid solutions were put to the test. As described above, the users had to evaluate and compare certain video clips for quality or synchronicity or mark screenshots for position of the signer video. (This took 25 minutes).

Finally, an interview was conducted with both free and multiple choice questions on general aspects of a sign language interpreter for digital TV. (This took 10 minutes)

4.3 User profile

The ten test users formed a heterogeneous group. The six women and four men were aged between 16 and 65 years. Half of the testers were deaf from birth and therefore the sign language background of the users was very different. Almost half used sign language as children, most of the others learned and used sign language at school only and one tester has been using sign language since the age of 27.

Concerning the users' media usage, the "Pre-task" questionnaire opening each individual user session showed that all 10 users own a PC and use it almost every day. While all use it for emailing and 90

percent for Internet access, seven testers had also used it for video services. Of these, seven use it for watching genuine sign language videos and four for watching videos with sign language interpreter and one for subtitled DVDs. All testers have a TV set, seven use it daily and three several days a week. All of them receive digital television, namely DVB-T, -S and IPTV. Except for the IPTV viewer none of them uses a receiver with integrated broadband access. Nine of the ten testers watch broadcasts with an embedded sign language interpreter, although most of them find the occasions of use too rare, the signer size too small and therefore often switch to subtitles. Five of the testers gave a school mark "average" regarding the current presentation of sign language interpretation, three rated as "good" and two as "bad" and "worse than bad".

4.4   Evaluation

4.4.1 Part One: Task Result: Evaluation of HbbTV On demand Application

Task 2.1: Please watch the news programme with sign language interpreter.

Of all ten testers only one person needed the interviewer's help in launching the signer video, the other nine testers launched the signer video without any help. The one person who needed help was not able to understand the task at first, but succeeded in a second run.

Comments: One tester had difficulties in reading the text because of the subjectively small font size. One more person was quite unsettled, the task was not clear to her, but she succeeded without help in the end.

Question 2.1.a: What do you think about the duration of the text popup (too short, appropriate, too long)?

Multiple Choice (appropriate/too short/too long): Eight persons found the duration of the text appearance appropriate, only two found it too short.

Comments: Those two persons who did find it "too short" commented that if one has got used to it, the text duration would probably be appropriate. For one tester the text popup lasted almost too long. Additionally it was asked if there could be an automatic setting, so that one does not need to actively approve each time a sign language service is provided. Another tester would have liked the service to respond faster. One person commented that the text as such should be shorter, another one similarly that even though she had no problems with the text application, some deaf users might have difficulties as they are not so used to reading long texts.

Question 2.1.b: What do you think about the font size of this text popup?

Multiple Choice (too small, appropriate, too big): Eight persons preferred the font size at it is, only two found it too small.

Question 2.1.c: Which position do you prefer for having the text popup?

Multiple Choice (bottom/top): Nine persons preferred the text position at the bottom, only one preferred to have it at the top. Comments: Seven persons said the position at the bottom was very common for deaf people, familiar from subtitles. Two testers stated that at a top position text within the main video might still be acceptable. The chosen colours for the text were agreed on, despite the comment that more luminance is needed for a higher contrast and to better differentiate text from the background. Used phrasing would be better shortened in order to optimise general understanding.

2.2: General feedback on the sign-language interpreter application

Question 2.2.1: What did you like most about this sign-language interpreter application?

Generally, most people mentioned more than one feature as the "best feature"; therefore the statements are counted to highlight the strongpoints of the service.

- Six statements: *one can start the signer service simply by one click.*
- Four statements: *signer image/ design*: "it's not as small as usual" and "we can see entire upper part of interpreter's body.
- Three statements: *Service in general:* "looked better than the commonly known sign language enhanced "Tagesschau" broadcast on channel Phoenix".
- Three statements*:* "free personal choice" *everyone can decide for themselves* whether they want to use a signer video or subtitles.

Question 2.2.2: What did you like least about this sign-language interpreter application?

Again the statements on the "worst points" are counted as there were multiple statements from the respective persons:

- Three statements (from older persons): the sign *interpreter was too fast* to understand even though the interpreter took down their arms for breaks between the news topics they said they missed her doing this. These testers would have preferred a summarising form of translation rather than the exact and fast approach that had been chosen by the interpreter.
- Three statements on the *background colour of the interpreter video.* Wish for more contrast for the interpreter's hands through a darker background
- Three statements *on the size of the sign language*. Two of these stated "should be smaller", one "is too small".
- Two statements: *quality of sign interpreter video: "Too blurry".*
- One statement: Service *should offer configuration* options for the user (background, position, size, and transparency).

Summary:

For the majority of the testers it was easy to start the signer service, the general approach was welcomed very positively and basic usability seems to be granted. Basic characteristics like image size, font size and the position of the popup text were by the majority positively rated and can be recommended. The strongpoints of the service are clearly *easy access* the service ("just one click") and the option to choose the signer "on demand" *(free choice).*

However, to improve the service, the test results suggest the following points:

- As little text as possible! A consideration to be explored could be to better choose symbols instead of text ("red hand" icon on the screen which can be found on a remote control, for example).
- If there are texts these should be in very simple language.
- The letters could be slightly bigger than in our implementation as some users found them too small.
- Strong contrast in the background of the signing video to bring out the interpreter's hands is important.
- Configuration options for each user for changing the size and the position of the signer are a wanted feature (also implicitly shown through differing judgements on the size of signing video).

Finally, one of the weak points was (for some of the elderly testers) the mode of our quite fast complete signing translation. This however is not a feature of the application / service we tested but rather a comment on the type of translation. Some of the younger testers commented that they want a "literal" and complete full translation.

4.4.2 Part Two: Tasks and questions for evaluating valid parameters for future hybrid sign language interpreter solutions

1. Synchronisation

Task 3.1.1: You are watching a video clip from Tagesschau enhanced with a sign language interpreter video. Please give a sign when you lose context.

> *Background*: For this part of the test the signing video is delayed compared to main programme. *Starting point is the usual human factor-based delay (mental processing and executing gestures) of sign language output regarding the main input and their according video presentation. The question here was: Up to which point is the delay acceptable for users? In order to find out about this acceptance threshold a video sample was prepared, starting a synchronous ("real world"-delay see above) delivery of main and sign language video, and then introducing an increasing delay of the sign language video.*

Five testers did not recognise a loss of context at all and were not irritated. Two gave a sign after 40 seconds and three at 25, 23 and 22 seconds.

Comments: Three of the testers stated that they did not feel irritated at all while two more stated they did not recognise a synchronicity problem here. One person mentioned that she is always watching the mostly the signer and not really the TV broadcast and therefore is not able to see differences.

Task 3.1.2: **It is possible to have the main and the sign language videos completely and simultaneously synchronised, thus eliminating the "real world time lag" of sign language translation. Please watch the video sample. How do you find the absolutely simultaneous presentation (good, bad)?**

> *Background*: Signing video is exactly simultaneous to main broadcast video, *i.e., even the real world delay is not there anymore. The approach is to technically synchronise both videos and remove all delay, so that sign language sentences fit exactly to the main video speech. Here, both videos were synchronised manually and re-encoded, so the test person had to judge if this approach eases the understanding.*

All ten of the testers rated the sample as good.

Comments: Nine persons judged the scenario explicitly as good in their free comments, one stating that the synchronicity was "just perfect" and two adding that it is an uncommon, difficult and strange experience but however enjoyable. One more person was not able to recognise a difference to the preceding test scenario.

Task 3.1.3: You are watching a video clip from Tagesschau enhanced with a sign language interpreter video. Please give a sign when you lose context.

Background: Signing video is ahead of main broadcast. *Here, the question is, will the acceptance increase if the user is able to access the language earlier than the main video? This third variant was to technically delay the main video, so that the sign language video is ahead. In the future this might be the case e.g. if a sign language video will be streamed via IP and added to the broadcast main video at the end device, the IP stream will be several seconds ahead of the encoding and multiplexing induced broadcast delay. To find out the acceptance threshold a video sample was prepared, starting a synchronous play of main and sign language video, and then introducing an increasing delay of the sign language video*

Six testers gave a sign indicating loss of context around the cut from the topic sport to the weather forecast at 1:12, 1:05, 1:04, 1:02, 0:55 and 0:52 minutes. Three persons did not recognise a loss of context and did not give a sign. Another person recognised a loss of context at 0:40 minutes.

Comments: Several testers recognised the conflict between the sports news topic still running while the sign language interpreter had already started to translate the weather forecast. One person found the idea of having the signer ahead attractive, as knowing the next topic beforehand makes it easier for her to focus on the main video.
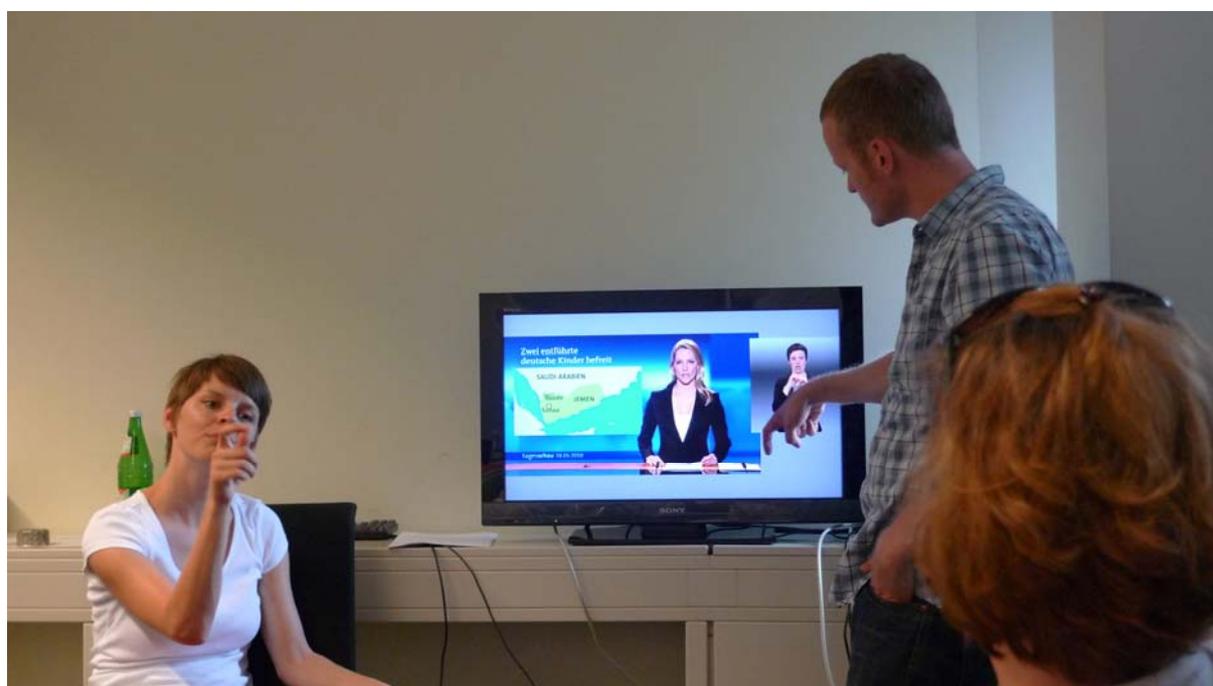
Figure 10. An expert test at rbb

Summary:

Looking at the results two main conclusions which might seem surprising arise:

1. The manually synchronised solution with the exact simultaneous sign translation is the clear winner with all of the testers judging "good" and nine explicitly positive comments for a new and unfamiliar service.

2. The synchronisation delay (starting from the natural delay) that is due to technological reasons does not seem to be as dramatic as expected. Only half of the testers found the increasing delay irritating and in six free comments it was voiced explicitly "I did not see a problem". However, values for the buffering to reach optimal results were gained (see below).

In more detail: Although the delayed signer video (first scenario) irritated users generally less than the "ahead" example (third scenario), the acceptance threshold in the case of the ahead running signer video is much higher in comparison. The delayed signer appears as a sound solution as six testers did not recognise a loss of context here, however, the rest of the testers were confused quite early, at around 20 seconds. In the "signer ahead" scenario people recognised a conflict much later. This scenario irritated more people but considerably later, therefore this approach can be called sound too, but with different effects. Generally, delayed sign language might have reached better results because it is well known to the users, while a signer running slightly ahead might have appeared unfamiliar.

For a future implementation it is crucial to build in a video buffer of a certain size, in order to realise each of the given solutions. To tackle human based delay and network-induced buffering (*signer delayed*), to totally *synchronise signer* and main video with different time codes and sources and to manage the growing delay of DVB broadcasts (High Definition and Standard Definition video encoding, multiplexing) compared to a real-time IP-access of a video stream (*signer running ahead*) buffering on the target device is needed.

For Germany a typical end-to-end delay of digital broadcast is up to 3-4 seconds for DVB-S. IPTV operators which use DVB-S as their source add 5+ seconds delay on top. This means that a receiver cannot just play the video streams as they arrive as this would not meet the expectations of users. In a future hybrid delivery of such a service a mechanism has to be used, so a receiver is able to resynchronize both service components for presentation. For sign language the presentation of broadcast video and Internet video need not be frame accurate. Buffer sizes for Internet and/or broadcast delivered content must be designed for the targeted combination of delivery networks with their specific end-to-end delays.

The results of the test show that a buffer size ensuring the limit of a maximum delay of approximately 1.76 seconds seems to be the best solution for a delayed signer approach. Depending on the absolute and real delay the main video must be buffered in order to gain keep the delay within this limit.

For a signer ahead approach, a buffer size ensuring the limit of a maximum remaining delay of approximately 4.16 seconds seems to be the best solution. Depending on the absolute/real delay the signer video must be buffered in order to keep the delay within the calculated limit.

> (Min. duration in test) 22 sec * (fps difference) 2 fps / (Standard fps) 25fps = <u>1.76 sec</u>

> (Min. duration in test) 52 sec * (fps difference) 2 fps / (Standard fps) 25fps = <u>4.16 sec</u>

1. Video Bitrate

Task 3.2.1: You are watching a video clip from Tagesschau enhanced with a sign language interpreter video. Please rate the image quality with a school mark (1="very good", 2="good", 3="satisfying", 4="bad", 5="very bad")

> *Background: Low video bitrate at 200kbit/s*

Multiple Choice: Seven testers judged the image quality of the low bitrate sample as "good". "Very good", "satisfying" and "bad" got one vote each.

Comments: Two persons stated that they could not see any compression artefacts or smearing. One found the image a little bit blurred.

Task 3.2.2: You are watching a video clip from Tagesschau enhanced with a sign language interpreter video. Please rate the image quality with a school mark (1="very good", 2="good", 3="satisfying", 4="bad", 5="very bad")

> *Background: Medium video bitrate at 300kbit/s*

Multiple Choices: Seven testers judged the image quality of the medium bitrate sample as "good". Three persons rated the example as "satisfying".

Comments: Three testers stated that they could not see any difference to the first example. Two persons found the image sharper and the colours more pronounced. One person judged the image as having a lower quality than the first sample, one more found the image sometimes smeary.

Task 3.2.3: You are watching a video clip from Tagesschau enhanced with a sign language interpreter video. Please rate the image quality with a school mark (1="very good", 2="good", 3="satisfying", 4="bad", 5="very bad")

> *Background: High video bitrate at 450kbit/s*

Multiple Choices: Six testers judged the image quality of the medium bitrate sample as "good". Two persons rated the example as "very good", 1 as ""satisfying". And another one rated the quality as "very bad".

Comments: Five testers stated that they cannot see any difference to the first two examples. One more person found the image now clearer.

Summary:

Medium bitrate of 300kbit/s was ahead by a nose. But in summary both, the results and the comments of the testers showed clearly that they could hardly identify differences between the three versions.

3. Image Size

Small image size 208x204 pixels

Task 3.3.1: You are watching a video clip from Tagesschau enhanced with a sign language interpreter video. Please rate the image size with a school mark (1="very good", 2="good", 3="satisfying", 4="bad", 5="very bad")

Background: Small image size 208x204 pixels

	In percent				
User Rating	1 Very good	2 Good	3 Satisfying	4 Bad	5 Very bad
Rate the image size	*0*	*20*	*40*	*20*	*20*

Comments: Seven persons stated that the size of the signer video was too small and that one "could hardly follow the movements of hands and fingers and was therefore not able to watch the main video too". Only one tester found the image size explicitly enjoyable.

Task 3.3.2: You are watching a video clip from Tagesschau enhanced with a sign language interpreter video. Please rate the image size with a school mark (1="very good", 2="good", 3="satisfying", 4="bad", 5="very bad")

Background: Medium image size 277x272 pixels

	In percent				
User Rating	1 Very good	2 Good	3 Satisfying	4 Bad	5 Very bad
Rate the image size	*30*	*60*	*10*	0	0

Comments: Three testers stated the image size here is optimal, one person found it enjoyable. One tester would prefer an even bigger size while another one thought that this image size is too big, "between small and medium would be good".

Task 3.3.3: You are watching a video clip from Tagesschau enhanced with a sign language interpreter video. Please rate the image size with a school mark (1="very good", 2="good", 3="satisfying", 4="bad", 5="very bad")

Background: Large image size 416x402 pixels

	In percent				
User Rating	1 Very good	2 Good	3 Satisfying	4 Bad	5 Very bad
Rate the image size	*40*	*10*	*30*	10	10

Comments: Four testers found the image size is too big here, three persons stated that image size is optimal for them. Three more persons found the image size good, one of them observed that the main video is now difficult to follow.

Summary:

The values for the marks from "very good to satisfying" for "positive evaluation" can be summed up and compared to the sum of the values for "bad and very bad" in order to get a more pointed evaluation. Doing this, the medium size image is the clear winner (100% positive), especially given the fact that it got no bad marks at all and is such a non-controversial solution.

The large image comes second best. However, this is controversial with 20 percent of the users judging "bad" or "very bad". The small option was clearly unacceptable with 40 percent giving a negative evaluation.

4. Positioning

Task 3.4.1: Please rate the position of the singer on the screen with a school mark (1="very good", 2="good", 3="satisfying", 4="bad", 5="very bad")

User Rating / Signer position	In percent					
	1 Very good	2 Good	3 Satisfying	4 Bad	5 Very bad	
	Acceptable			*Not acceptable*		
Right top	**0**	**10**	**70**	**20**	**0**	Overlapping
Right middle	0	0	60	40	0	
Right bottom	10	10	40	40	20	
Left top	0	0	10	70	20	
Left middle	0	0	20	60	20	
Left bottom	0	0	10	60	30	
Right top	10	30	30	20	10	Adjoining
Right middle	**10**	**30**	**40**	**20**	**0**	
Right bottom	0	30	40	30	0	
Left top	0	0	20	50	30	
Left middle	0	0	30	50	20	
Left bottom	0	0	20	30	50	

Within the overlapping examples "right top" got the highest votes: 80 percent within the "acceptable" range and 20 percent within "not acceptable", which leads to an overall acceptance rate of 60 percent. Within the adjoining examples nearly the same vote was reached with an overall acceptance of 60 percent (delta value) for "right middle": 80 percent within "acceptable" and 20 percent within "not acceptable" range.

User rating	In percent		
	Adjoining	Light Overlapping	High overlapping
What do you prefer most?	**30**	**50**	10

Half of the users prefer a "light overlapping", 30 percent prefer "adjoining". "High overlapping" is desired by one only user.

Summary:

The testers are used to having a sign language interpreter on the right hand side. This seems to have influenced the results of this task: the high acceptance of the right aligned examples. Light overlapping between signer and main video was preferred over an adjoining presentation, which leads to the result of having the signer video preferably at right top position.

4.4.3 Part Three: Results: Questions on general aspects of a sign language interpreter for digital TV

Question 4.1: *Would you generally use such novel sign language interpreter services on TV?*

All testers said, yes they would use this kind of interpreter service on TV.

Comments (three statements from three different persons):

- It is great to choose a service individually, either the signer video or subtitles
- The use of signer video services depends on the nature of the programme as well as on the individual sign language interpreter and their abilities. For watching magazines, documentaries and news shows a signer service is preferred.
- The handling has to be simple; therefore the menus have to be easy, navigation paths short and of a simple structure.

Summary:

The fact that all users said that they would use such a service even though only a prototype version with a yet unfamiliar handling was tested, underlines the great demand in this target group for signing services on TV and also the testers' openness to novel types of services. The testers emphasize the desire for personal customization and quantity; the usability should be very simple, like in the tested service.

Question 4.2: *How important is it for you to adjust position and size of the signer video? (The testers could choose between three answers "very important", "important" and "not important")*

Multiple Choices: Eight persons rated this option as "very important", two as "important".

Comments:

- Three statements: It is important to adjust settings based on individual needs.
- Two statements: It is not so important to adjust settings based on individual needs, however, one of the persons making this statement added that the sign translation service has to be on the right side and should be available in maybe three different sizes to select from. Consequently this person does want individual settings. This person stressed that not too many options should be available.
- One statement: The users would probably get used to a predetermined solution.

Summary:

The fact that all users rated the option to adjust the position and size of the signer video based on individual needs as "very important" or else as "important" shows that users definitely want such a feature. The free comments underline this with four statements explicitly demanding this option. However one person suggests that there should be a limited number of options to choose from in order not to make things too complicated. Therefore, a compromise solution in terms of position and size could be a fixed position on the right hand side with three sizes to select from.

Question 4.3: What approach for handling synchronicity do you like best (signer delayed, synchronous, running ahead)?

Multiple Choices: Surprisingly no one voted for the (naturally) delayed signer modelled on the "real world delay". Seven testers explicitly preferred the manually created simultaneous solution and, three people the signer running ahead of the main video.

Comments:

- Two statements: The synchronicity of both videos is explicitly preferred.

- Two statements: It is better when the signer video is running ahead because it is easier to follow the main video.

- One statement: The signer video would be focused on firstly and the main video secondly.

- One statement: A summary would be better instead of a literal translation.

Summary:

The results of this question nicely underline the outcomes of the respective tasks described above which had shown overwhelming support for the simultaneous option. This approach should be pursued and, obviously a possible deviation would rather be accepted if the signer video was running ahead then if it were delayed. The fact that no users rated the naturally delayed signer as "best" is surprising, as the target group is used to this mode. The justification provided by the testers, that a synchronised or running ahead solution allows the user to better follow the main video, makes sense.

Question 4.4: Do you have any final comments for the accessible design of a new signer video TV-service?

- Four statements: The image size of signer video could be bigger than in our chosen example.
- Four Statements on the background:
  - Two persons complained that the background to the sign language interpreter was a gradient background; one of them also wanted a brighter non-gradient background.
  - One person liked the gradient background.
  - Another person stated that different background colours for the sign language interpreter should be tested.
- Two statements: A perspectively reduced (30 degrees) signer video for a three-dimensional effect would be better.
- One statement: The interpreter should be dressed in black.
- One statement: The exact simultaneous synchronicity of signer and main video was particularly important for one person.
- One statement: Animated avatars should not be used as translators, because of missing visemes and facial expressions.

Even though the question was targeted at learning about the actual design and look and feel of signing solutions some persons remarked on general issues.

- Three statements: Three persons wished for more programmes with sign language interpreters in general, especially for information programmes such as news and discussions while they found the sign language support for movies not as important.
- Three statements: the sign language used in this test was too fast.
- Two statements: adjust settings oneself (video size and position).
- Two statements: it is very positive if users can decide for themselves whether one wants to use subtitles or a signer video, users want to have both services on offer.

4.5 Conclusion

Detailed and concrete recommendations on the single aspects of presentation and technological set-up of sign language solutions in digital TV will be extracted from the above test results and issued in the finalising deliverables of DTV4All.

Summarising the results of part one of this test, the novel HbbTV-video on demand solution was very much welcomed by the testers which shows that people are open to new approaches if they make sense to them. Concerning usability the presented solution was generally satisfying, it could however be improved in some smaller aspects. People want a service that is easy to understand and offers good usability.

Concerning synchronisation issues between the signer and the main video, the testers surprisingly welcomed a manually established total simultaneous synchronicity which overcomes even the real world delay. For reaching a high acceptance of de-synchronous videos a buffer size is needed which assures that offset limits of 1.76 seconds for a delayed signer and 4.16 seconds for signer ahead are not exceeded.

When it comes to video bitrates, the results show that the usage of modern state-of-the-art video codecs like H.264 can offer both an optimal image quality and small bandwidth for a signer video.

Speaking of the position of the signer video, people expect it to be on the right hand side of the screen, whether it is located at the top or in the middle. A light overlapping between both videos can save some space on the screen and guarantees sufficient image sizes without obscuring important information.

This test emphasized the desire for personal customization in order to fit the individual needs of deaf people, especially their individual seeing abilities. This needs to be addressed by the provision of possibilities for adapting the video image size and position, which is supported by 100 percent of the test group members.

5. Appendices (further Questionnaires etc.)

5.1 Subjective evaluation of the enhanced AD emerging services report

<div align="center">

Subjective assessment

Emerging accessibility of services:

Audio Description on digital TV systems

Report on the results

</div>

<div align="center">

Anna Vilaró Soler

Aitor Rodríguez Alsina

Dept. Enginyeria TVC

July 2010

</div>

Motivation

This document presents the results and conclusions regarding the evaluation tests that were carried out within the framework of the DTV4All project on Emerging Internet services: audio description for digital television.

Each of the different scenarios presented is a different service with distinct features, so the results are presented separately for each of the five scenarios evaluated. However, prior to the presentation of the results and conclusions, details of the methodology and procedures for data collection are given. The document structure is the following:

Document contents

Section 1   Methodology

1.1.   Show participants

The selection of participants followed the recommendations set by the International Telecommunications Union (ITU) on methodologies for subjective quality assessment of television systems (ITU-R BT.500-11). We have worked with a sample of participants not experts, as they neither work in the area quality in multimedia systems nor are consultants in this field. The test participation was voluntary. The sample of participants was 14 people (6 women and 8 men). The ages of the participants were between 21 and 40 years, with an average age of 28.8 years. They all understood and spoke Catalan well and were computer users on a daily basis.

Since the completion of testing of an AD scenario means an interaction with the applications of AD services. For the quality evaluation of the AD audio compared to the image we have established criteria for selecting test participants. With the aim of monitoring the homogeneity of the participants it was found that they had no problems of perception. Perception was assessed through vision and hearing tests. These perception tests and their results are briefly described below. With respect to the visual tests, visual acuity was assessed with a computerized version of the Landolt test (Figure 1). All participants had normal visual acuity, which ensures that the participants were able to read the menu texts, evaluate the perceived image quality received (for example, by defining image details), read the test instructions and questionnaires, etc. We also checked the possible existence of colour blindness with a printed version of the Ishihara plates (Figure 2). Among the participants we did not detect any anomalies associated with colour blindness. Concerning auditory perception, participants answered a questionnaire about everyday situations indicative of potential hearing loss. None of the participants reported hearing problems that could put into question the performance of the tests.

Figure 1: Landolt test. This shows a "letter C" of a certain size and participants must indicate in which direction the opening is presented. In each trial the size of the stimulus is reduced. The size for which openness is no longer perceived indicates the degree of visual acuity.
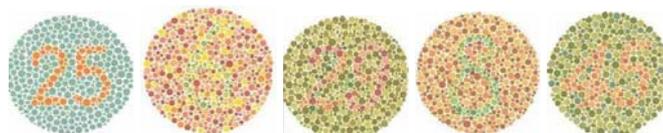
Figure 2: Ishihara plates. These present a film in which figures repeatedly appear and each time and the participant must indicate which figure appeared. Depending on whether you have some type of colour blindness or not, will receive a particular figure or another, or none.

Additionally, participants were surveyed with the aim of defining a profile of the sample of users in relation to the habits and preferences relating to audiovisual content. Presented below are the participant details.

In relation to the degree of prior knowledge of audio description services, most of the participants did indicate that they knew what audio description is and would be able to explain it (Figure 3). Regarding its use, none of the participants indicated having previously used an audio description service.
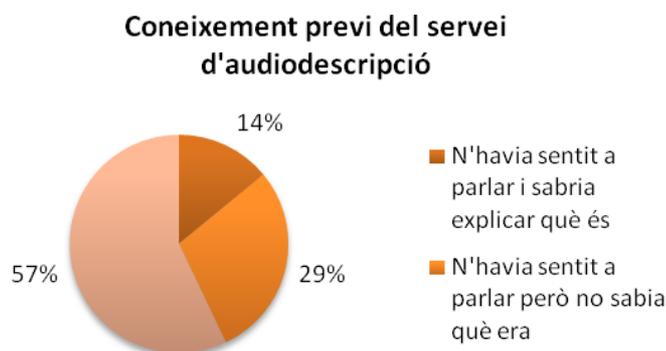

Figure 3: Prior knowledge of audio description services

On the question regarding the use of audiovisual content via the Internet, about a quarter of the participants used audiovisual material daily, whether on work days or holidays. Moreover, almost half of the participants stated that they consumed audiovisual material during weekdays. Approximately one quarter of the participants consumed audiovisual material over the Internet sporadically or never (Figure 4).
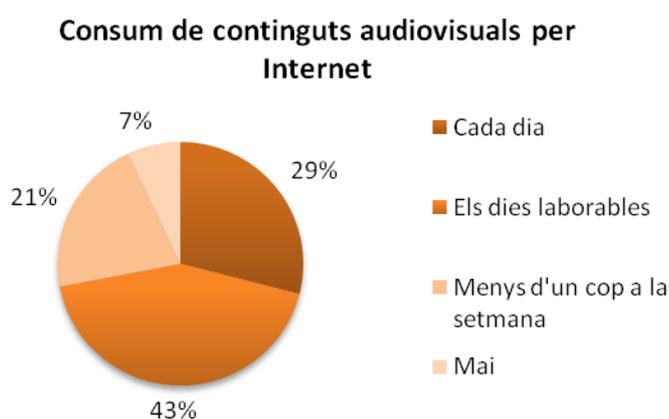
Figure 4: Reception of audiovisual content over the Internet

Some of the other matters related to the profile of the participants are presented below. Regarding reception of content through the portal 3alacarta (TV on demand), 57% of participants reported that they had used it, although none of them had used it on a weekly basis. The remaining participants reported never having used it. When asked "Would you like audiovisual content for its TV schedule when you cannot see it on a television?" 56% of participants indicated they would very much like it, or 50/50 (28%, respectively), while the remaining 46% of participants indicated that they were

indifferent to such provision. Finally, when asked what device they would use for reception of the audiovisual content if they could download it, 78% of participants reported a laptop, while only 14% said they preferred on a television. The remaining 8% opted for other options (mobile phone or desktop PC).

1.2. Data collection

The evaluation of various scenarios involving a time limit tied to the live broadcast content was organized with the aim of optimizing the duration of each session, which consisted among other things of two participants per session. Figure 5 shows the distribution of time devoted to different scenarios during a session.

	Broadcasting "Riera" TV3								
Hours	15.55 - 16.05	16.05 – 16.15	16.15 – 16.20	16.20 - 16.30	16.30 – 16.40	16.40 – 16.45	16.45 – 16.55	16.55 – 17.05	17.05 – 17.15
Participant A	Scenario 1 AD TDT	Scenario 1 AD STREAMING	Scenario 1 QUESTIONNAIRE	Scenario 2 AD TDT (AURIC)	Scenario 2 QUESTIONNAIRE		Scenario 3.1 DOWNLOAD TV QUESTIONNAIRE	Scenario 3.2 DOWNLOAD PC QUESTIONNAIRE	Scenario 3.3 DOWNLOAD PODCAST QUESTIONNAIRE
Participant B	Scenario 2 AD DTT (Auric)	Scenario 2 QUESTIONNAIRE		Scenario 1 AD TDT	Scenario 1 AD STREAMING	Scenario 1 QUESTIONNAIRE	Scenario 3.1 DOWNLOAD TV QUESTIONNAIRE	Scenario 3.2 DOWNLOAD PC QUESTIONNAIRE	Scenario 3.3 DOWNLOAD PODCAST QUESTIONNAIRE

Figure 5: Time distribution for each stage in a test session

As shown, each participant performed the tests in a different order, so the two participants in a session could share some period of display, thereby shortening the total duration of the session. The sessions conducted lasted about an hour and fifteen minutes.

To design protocols to inform participants, written documents, and instructions were created, and specific verbal instructions were drafted. Participants agreed to collaborate and had a certain amount of information related to the objectives of the test obtained during the request for participation, as requested by the ethics protocol (Annex II). On arrival at the laboratory participants were properly informed about the objectives and tasks to be performed through a written document. After any possible doubts were resolved, and having verified that the participant had correctly understood the purpose and its participation, they were asked to sign the consent to participation form in order to follow ethics protocols (Annex III). After this procedure tests started. Test managers were present at all times during testing sessions.

Regarding the procedure followed for the assessment of each scenario, we followed the previously established sequence. It was important to assess scenarios and the services associated with them that participants were used to. The standardized way of arranging this is to define a task representative of the service to be evaluated which the participant must perform. Thus, in order to obtain the corresponding subjective assessment, participants were asked to perform a specific task that was performed for various emerging services offered (e.g. tune specific audiovisual content in a particular system). For each service, it was assessed whether the participants had read a printed sheet that had a description of the work to be carried out on it (Annex IV), and then the participant was asked to provide an oral summary to verify that they had understood correctly what was written on the sheet. Then the task was carried out by the participant. This task could be, for example, "tune the channel number 120, La Riera with audio description via a live streaming system". In all cases, it was specified that the task was to carefully observe the quality of the audiovisual material that was received, both in terms of its auditory and visual aspects. After the task, the questionnaire was administered. The questionnaires were delivered to participants with ample time for a comprehensive reply. Each questionnaire contained a cover sheet explaining its purpose and test specifications. Once the participants had rated all the services (Figure 5) they were informed that the test had finished. Finally, the participants were thanked for their collaboration.

1.3. Measuring Instruments

The instruments that were used to measure subjective evaluation of emerging services accessibility provided by TV3 correspond to a set of five questionnaires, each of which had been adapted to a specific scenario by adding to or taking away specific questions.

The aim of these questionnaires was to gather, in an organized way, the indicators of the psychological variables involved in the tests. That is, the questionnaires helped us to obtain information from participants about their attitudes and opinions on a particular aspect to be assessed which could not be measured directly. In this sense, the questionnaires used had three stages, which assessed three aspects related to emerging services accessibility: Usability, Utility and Quality of the received signal. Therefore the questionnaires were structured explicitly in different parts, corresponding to these three issues.

The types of questions used were drafted to answer questions with closed-type Likert scale consisting of seven response alternatives from "strongly agree" to "strongly disagree" (Figure 6). The use of Likert scales implies that all questions concerning a particular psychological variable (e.g. perceived quality) are monotonically linked to this variable. In other words, the questions always used a one-dimensional scale to assess some issues. Hence, systematic variations in the responses of people to questions of scale are due solely to differences between people.

| Strongly agree | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Strongly disagree |
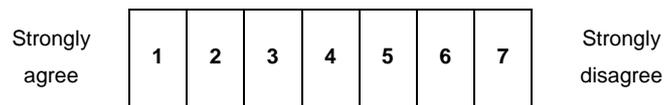
Figure 6: Scale of response used in the questionnaires for the Likert scale

Moreover, the questionnaires also included open-answer questions concerning the assessment of positive and negative aspects of the service assessed, as well as aspects that may have been missing. The idea was to collect the impressions of the participants beyond closed questions which otherwise would not be possible to gather. These open questions are part of the section to assess the usefulness of the proposed service.

Through questionnaires, comments from participants were also collected. To facilitate the participants expressing their opinions a space was reserved for comments below each question in the questionnaire which the participant filled if they wanted to provide more information.

In presenting results for each scenario contained herein, the data are structured in three parts, corresponding to the constituent parts of the questionnaires. Thus, the relation between the questionnaires and the type of questions is as follows:

- Usability: complementary observations and numerical answers.

- Utility: numerical responses, open responses and additional comments.

- Quality: Complementary observations and numerical answers.

For specific questions about the variable usability, the questionnaire "Post-study system usability questionnaire" (Lewis, 1995) was adapted. The usability scale, in this case consists of two subscales. On the one hand, the scale of ease of use, on the other hand, the scale refers to the so-called interface elements used to interact with the system (which includes the use of elements, graphics, and language). For other variables, we raised questions that meet the criteria of reliability and validity. Compliance with these criteria has been reviewed by two observers (for example, all questions of the questionnaires are formulated in a positive sense, to avoid problems of comprehension).

Finally, it is important to remember that the questions of the questionnaires refer to the tasks performed by the participant in the respective scenarios.

1.4.   General considerations

Before proceeding to analyze in detail the different scenarios we have considered it appropriate to introduce a section on general considerations that includes aspects that are common to all scenarios.

The data presented below correspond to the assessment made by 14 participants on five different scenarios. Depending on the complexity of the scenarios, the questionnaires contained a variable number of questions.

Regardless of the procedure followed in the analysis, once collected the data was entered into a spreadsheet for further analysis. This process includes data for assessing the Likert scale, open

questions and observations that could be optional to the participants. The information was gathered depending on the type of data in question.

In the case of numerical data corresponding to the Likert scale, data are presented after having been normalized. The use of standardized scores facilitates understanding, and allows comparisons between items that originally shared the same assessment scale. These scores are expressed on a scale of 0 to 10, where 0 indicates "strongly disagree" and 10 indicates "completely agree". The calculation of the normalized score was determined from the average values for each question. This value was obtained by reversing the scale of 1 to 7 and turning it into a 0 to 10 scale according to the calculation:

value_normalized = (7 - value_scale_Likert) * 10 / 6

Regarding the treatment of numerical data in special cases, e.g., where the items had missing values, the average was calculated taking into account the total number of responses. However, in the treatment of numeric data other descriptive indices were used such as frequency counts of individual assessments.

With respect to the analysis of open questions, we analyzed the content of the responses through the categorization concepts. The results of the open questions are outlined in this document as a collection of opinions.

Section 2.   Analysis of results

2.1.   Scenario 1. Live streaming Internet TV

In this scenario the audiovisual content was transmitted to the digital TV with audio description by live broadcast over the Internet. The contents were transmitted from TVC servers to the UAB laboratory.

a. Assessment of usability

To evaluate the usability, the questionnaire consisted of 12 Likert scale questions. As mentioned above, the variable usability is subdivided into two other variables: ease of use and the interface. The scale provided for ease of use covers question number 1 to question number 7. Moreover, the interface level is assessed with items ranging from question number 8 to question number 13. The questions regarding usability are presented in the table below.

USA_01	Overall, I am satisfied with the usability of this system
USA_02	This system is simple to use
USA_03	I can easily complete the tasks that I have been requested to do
USA_04	I can quickly complete tasks that I have been requested to do
USA_05	I can efficiently complete the tasks that I have been requested to do
USA_06	I feel comfortable using this system
USA_07	This system is easy to learn to use
USA_08	The information provided (text on the screen help feature) is clear
USA_09	The organization of information in the system screens is clear
USA_10	Navigation through the menu screen is simple
USA_11	The information provided has been effective in helping me complete the tasks
USA_12	The interface of this system is pleasant

Table 1: Numerical scale usability questions for scenario 1

Below are the results for each of the items in the usability scale (Figure 7). The values are mean normalized on a scale of 0 to 10, where 0 indicates "strongly disagree" and the value 10 indicates "totally agree". The colours refer to the subscale to which each question belongs.
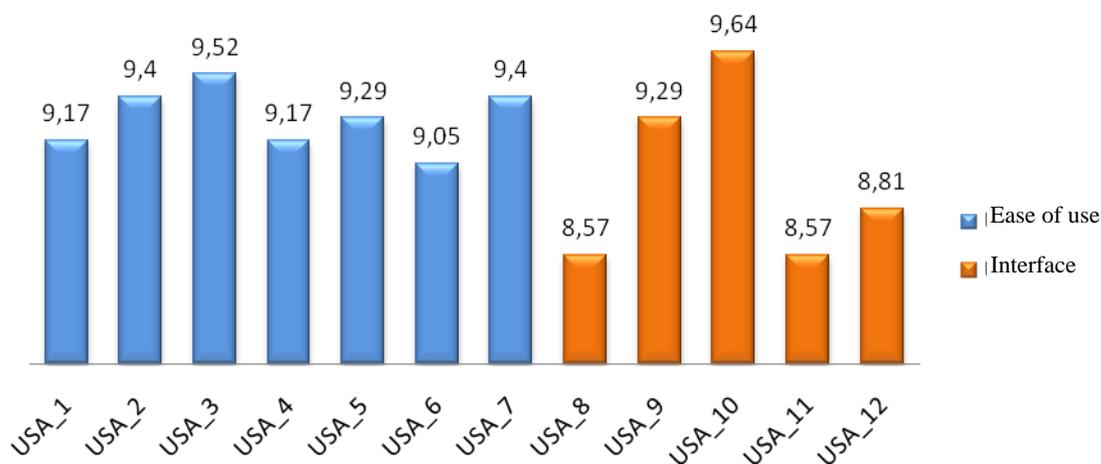
Figure 7: Normalized picture of the questions on usability for scenario 1

As shown by the graphic scores, all questions on usability provide values higher than 8.5 points, with an average score of 9.15 points for usability. The results indicate that, in general, participants rated very positively all aspects of the usability of the system for Live Internet TV streaming.

If we look more closely at the usability data, with special attention to the two subscales usability and interface shown in Figure 7, we can observe some differences in the scores. The mean for the subscale ease of use is 9.28 while the mean for the subscale interface is 8.9. With respect to the subscale ease of use we note that the scale scores are very similar for all the questions, whereas for the subscale interface, the values are more widely spread. Proof of this is that the maximum and minimum usability

scores are found in this subscale. It is interesting to note that the two lower scores refer to aspects of the textual information system. More thorough analysis of the data that makes up Figure 7 highlights the fact that there is little variability among participants in their responses to several questions, i.e., the data shows a significant degree of agreement among participants. This fact suggests that the scale presents a good level of assurance regarding the construct it is intended to measure.

Regarding the comments provided by participants about the usability of this system, reinforcing the numerical results are observations such as "is easier to activate by using the audio description." and "a major improvement in usability is that the menu system offers audio description." Also related to the navigation menus, reference was made to screen names which are technical names appearing in English, which may hinder the participants understanding.

b. Assessment of the utility

To assess the extent to which the Live Streaming Internet TV system was useful to participants, five Likert scale questions and 3 open questions were posed. The Likert scale questions and open questions are shown in the following tables.

UTI_13	This system has all the features and capabilities that I hope it would
UTI_14	In general, I think this service is useful
UTI_15	I have this service
UTI_16	If I had access to this service it would use it
UTI_20	Overall, I am satisfied with this system

Table 2: Numerical scale questions asked about the usability of scenario 1

UTI_17	What stands out as the main advantages of this service?
UTI_18	What stands out as the main disadvantages of this service?
UTI_19	If you were an audio description user, what other options you would like this service to offer?

Table 3: Questions about the utility of scenario1

Below are the results that evaluate the usefulness of the Live Streaming Internet TV system. Figure 8 shows the normalized average scores for each of the questions with numerical answers that have been presented previously. The value 0 indicates "strongly disagree" and the value 10 indicates "totally agree".
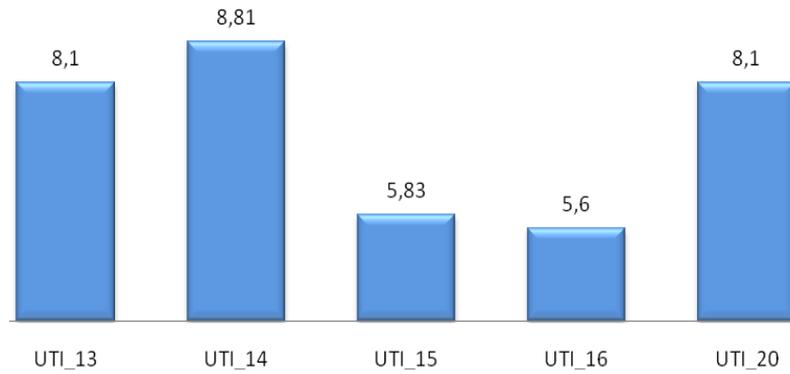
Figure 8: Normalized picture of the questions on the usefulness of scenario 1

The average overall utility score is 7.28. However, as shown in Figure 8, the results show an unbalanced assessment of the different items. If you look at the different scores, we can observe that questions 13, 14 and 20 obtained scores above 8, while questions 15 and 16 obtained lower scores, between 5.5 and 6.

In this sense we can say that users valued positively the usefulness of the system, including its functionality, and were satisfied with the system (questions 13, 14 and 20). Moreover, when the question of the questionnaire shows a personal appeal and asks about the hypothetical situation of having and using the system (questions 15 and 16), it seems that subjective judgments tend to rating the question "neither agree nor disagree". Here we examined more closely the data from questions 15 and 16 to ascertain the distribution of data that provided the averages shown in Figure 8. Thus, we found that for question 15 (would like to avail yourself of the service) there is no clear trend in the responses, so that all the possible answers are represented in the sample of participants (Figure 9). In addition, for question 16 (if you had access to the service would you use it), the distribution of participants' responses shows a tendency to be distributed among the most extreme responses "totally agree" and "strongly disagree" (Figure 10).

Figure 9: Frequency of responses to the question "I have this service"

Freqüència de respostes a la pregunta 16
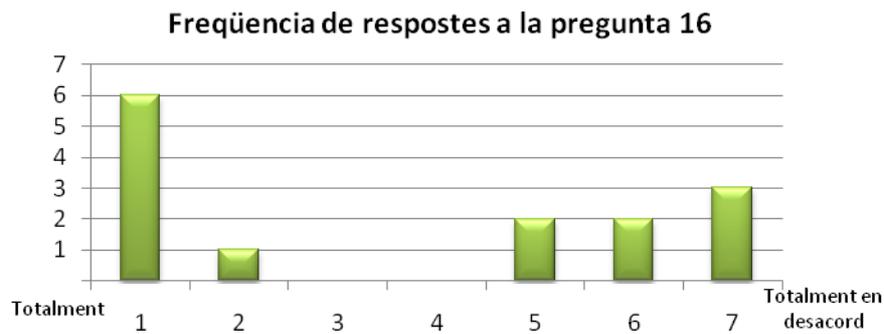
Figure 10: Frequency of responses to the question "If I had access to this service I would use it."

With respect to the open questions about the utility of the system, they were divided into positives, negatives and other options that should be included in the system.

Thus, as the main advantages of the system for the participants are the availability of content with high image quality and the usability of the system. For example, a contribution on quality says "lets me see the contents of TV3 with good quality." Regarding ease of use answers included "the clarity of command" or "is comfortable and easy to use." This last aspect has been highlighted by the majority of the participants. It is interesting to note that this broad agreement among participants about the ease of use correlated with the score on the subscale ease of use for the assessment of usability.

If we analyze the negative aspects of the system, we can group the contributions of the participants into two groups, both related to technical aspects. There is a need for an Internet connection that guarantees a constant bandwidth. Moreover, participants also mention the need to have (and having to learn to use) a set-top box to receive content. In relation to these aspects participants statements ranged from "we have the equipment and Internet connection" or "connections in this country do not apply to it."

In response to the question about what other options they would like this service to offer, participants identified several features to consider. We pooled the opinions collected. Participants emphasized the importance of the incorporation of information about the content provided, for example, the inclusion with audio description of a synopsis of the films on offer, or auditory information about the programming grid. One suggestion of particular interest is to include a specific sound at the beginning of a program to inform the user of the availability of audio description, i.e., that there is the possibility of activating the audio description service. Participants highlighted the potential value of the incorporation into the system of several multimedia options such as the ability to access web pages from the page system, a shortcut to view the schedule or the possibility of recording the contents of a programme.

c. Assessment of the quality

To evaluate the subjective quality of this scenario, the participants were asked four Likert scale questions and a question with a relative scale of 5 levels. The Likert scale questions are given below.

QUA_21	The image quality is good
QUA_22	The audio quality is good
QUA_23	After selecting the service, access to content is quick
QUA_24	There are no cuts or breaks

Table 4: Numerical scale questions about quality for scenario 1

The following are the results that evaluate the quality of the live streaming internet TV system. Figure 11 shows the normalized average scores for each of the questions above.
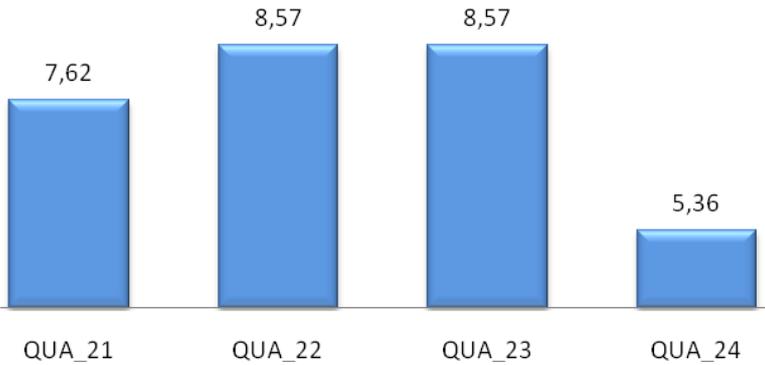

Figure 11: Normalized picture of the questions on quality for scenario 1.

As shown in Figure 11, in general the quality of the system was valued positively. However, these results show a negative result for image quality due mainly to cuts and breaks in the signal stream. This penalty can be seen in the results of question 24, where the average rating is 5.36. We also observed a lower score on question 21, which asked about the overall quality of the image. However, the scores suggest that these cuts or breaks were not critical enough to result in a bad user experience, or prevent the user from properly following the contents. On average, participants accepted the existence of cuts and breaks and assessed the image quality as being of a surprisingly high quality. The experience of the participants of Internet video technologies, which commonly display these small errors (usually due to network problems), may have resulted in them not penalizing too much the score for the image quality. Typical observations made about the image quality were: "There are some cuts, but not that difficult to understand" or "the image quality is good, but Freeview is better."

The subjective quality assessment of the audio, however, does not appear to have been penalized because of these cuts and jumps and the participants rated audio quality highly. Moreover, they also rated positively the speed with which they could access the service. In fact, they were very satisfied with it and were no comments made about this.

The question that asked the participants to compare the relative quality of the broadcast via streaming with the signal received by DTT was evaluated using a scale of 5 levels as follows.

QUA_25	To what extent is the overall quality degraded in the streaming system compared to the DTT system? 5. Imperceptible 4. Perceptible but not annoying 3. Slightly annoying 2. Annoying 1. Very annoying

Table 5: Question on a relative scale for scenario 1

Here are the results of question 25:
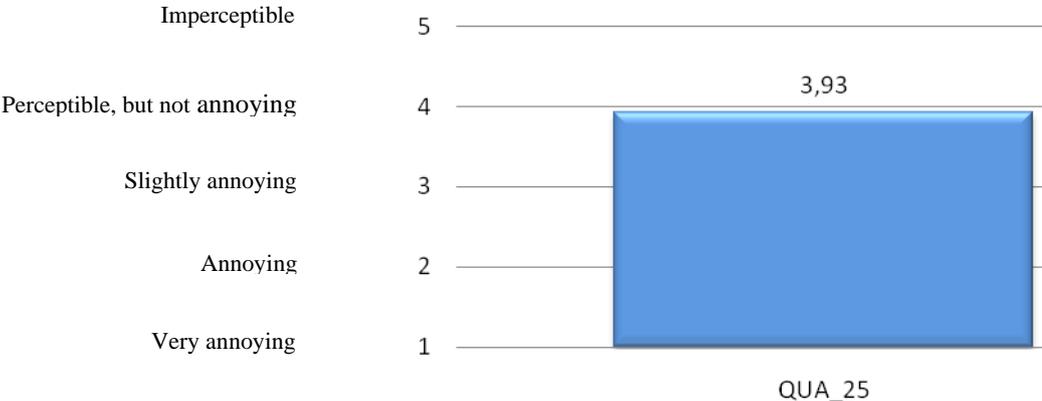
Figure 12: Normalized picture of the comparative question for scenario 1.
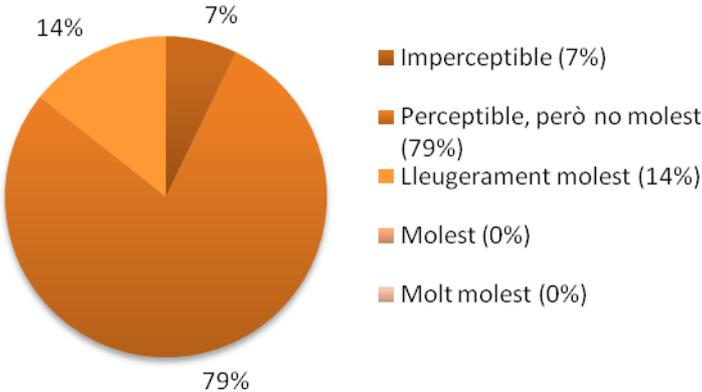
Figure 13: Frequency of responses to the question on comparative quality of scenario 1

The figures above show a high acceptance of the streaming system with respect to DTT although the differences in quality between the two systems have been clearly perceived. The vast majority of participants (79%) have rated this difference as "perceptible but not annoying." This allows you to think that, despite the presence of cuts and breaks in the image, participants found the quality of the

streaming service acceptable. None of the participants rated the quality of the content as "annoying" or "very annoying". This result is consistent with the data analyzed in the section satisfaction with the overall system utility.

d. Conclusions of the evaluation of scenario 1

In this scenario the system evaluated subjectively was live broadcast TV content over the Internet, also called Live Streaming Internet TV. The signal the system provided was compared with the signal provided by DTT to determine the difference in quality between the two systems. Below are summarized the results of the evaluation of this scenario according to the different aspects evaluated.

Aspect rated	Score
Usability	9,15
Ease of use	*9,28*
Graphical interface	*8,90*
Utility	7,28
Quality	7,53
Relative quality (question 25 score normalized out of 10)	*7,86*

Table 7: Summary of rating of scenario 1

The analysis of data from the questionnaires for this scenario indicates that, in general, the system has been rated very positively by the test participants. If we focus on the assessment of the different aspects that have been evaluated, we stress that usability has been the most valued aspect, both for the ease of use for the graphical user interface (GUI). Participants emphasized mainly the ease and speed of access to the Live Streaming Internet TV service as well as the simplicity of the GUI. Participants have suggested that the GUI could be improved with descriptions in Catalan of the main features of the scenario, as most do not know the meaning of the technical term 'Live Streaming' for example. Another suggestion was to introduce sound descriptions of the menus of the GUI to make it more accessible to people with visual disabilities.

The usefulness of the service has also been positively evaluated, especially with regard to the usefulness of the service for people with visual disabilities. But it seems that some participants do not place much importance on having the system at home and use the audio description service, while not necessary, either for them or any family member.

The assessment of the quality of this scenario has been affected by the presence of some specific cuts and breaks in the image stream from the signal, which results in difficulty understanding and monitoring the contents. In addition, users have rated the image quality significantly better on DTT than over the Internet but quantifying this difference as "perceptible but not annoying."

Although there are aspects of this scenario that can be improved, it has shown its feasibility and acceptance by users. Participants have found it very interesting being able to watch live online with a quality similar to what can be seen on DTT. According to these results, the quality of service with regard to cuts and breaks would be the main point to improve.

2.2. Scenario 2. AD reception: individual versus collective

The scenario of simultaneous reception of audiovisual contents with and without audio description proposes a solution for reconciling family television viewing habits in those homes with a member with visual disability. In order to take on board the tests, the participant had a laptop with Freeview tuner tuned to an AD channel which is received through headphones. The participant will see the image on the television (which in the home would be being viewed by other family members).

a. Appraisal of the utility

To estimate to which degree the system has seemed useful to the participants, the participants were asked 4 Likert scale questions and 3 open questions. In Tables 8 and 9 the questions are presented.

UTI_14	In general, I think this service is useful
UTI_15	I have this service
UTI_16	If I had access to this service I would use it
UTI_20	Overall, I am satisfied with this system

Table 8: Scenario 2 questions about utility on a numerical scale

UTI_17	What stands out as the main advantages of this service?
UTI_18	What stands out as the main disadvantages of this service?
UTI_19	If you were AD user, what other options would you like this service to offer?

Table 9: Questions about the usefulness of scenario 2

Next the results presented for the service. The bars represent the average scores normalized for each of the questions requiring numerical responses. Value 0 indicates "totally in disagreement", and value 10 indicates "totally in agreement".


Figure 14: Normalized scores for questions on the utility of scenario 2

The global average obtained for the utility of the system AD reception is 6.69. This score suggests that, in general, the utility of this service is appraised positively. If we take the scores for each question into account, we observe how the general utility (question 14) is the element best appraised. Also a very good appraisal is resulting from using the system; question 20 makes reference to user satisfaction with the system. However, when the participants were asked if they would like to have the service (question 15) and about if they would use the service (question 16) the responses are less positive with scores of 5.60 and 5.12 respectively.

We noticed the different degree of variability in the answers to each question. It is necessary to highlight that there is a great unanimity when responding to question 14. In the same way, in question 20 we find a general consensus among the participants with highest scores distributed between the values 2 and 3 on the Likert scale. On the other hand, however, and in the same way as for scenario 1, we saw a marked dispersion in the responses to questions 15 and 16. The data shows that when the participants are asked if they would like to have the service, the opinions are diverse and there is not any marked trend. When the participants are asked if they would use the service, it seems that there is a significant group within the participants that do have a clear idea about whether they would use the service or not (Figures 15 and 16).

Figure 15: Frequency of responses to the question "I have this service"

Figure 16: Frequency of responses to the question "If I had access to this service I would use it."

From the responses to the open questions designed to gather information on the advantages and disadvantages of this system, we have collected the following opinions. The principal strength related to use of AD, participants emphasized the improvement as the possibility to have audio description

individually. This is reflected in comments like "you can see the series together and choose whether or not you want to listen to audio description."

As for negative aspects, most participants stressed the fact that you have to have a Freeview receiver and a laptop computer. Some participants also pointed to potential difficulties in placing the laptop computer near the TV set (cable connecting to the battery, the receiving antenna, the length of headphone cable, etc.) to get AD in the presence of hearing family members. Also related to the use of a laptop, participants mentioned that greatly improved user experience could be achieved using a device much smaller than a laptop that could tune to the DTT signal (e.g. the size of a mobile phone).

b. Appraisal of the quality

To evaluate the subjective quality of this scenario, we asked participants five Likert scale questions and a question allowing responses on an absolute scale of 5 levels. The Likert scale questions are given below.

QUA_21	The audio quality is good in my headphones
QUA_22	The audio from my headset does not have cuts or breaks
QUA_23	The audio from my headphones and the audio from the television were exactly synchronized
QUA_24	There was no imbalance between the headset and TV
QUA_25	After selecting the service, access to content was quick

Taula 10: Numerical scale questions on quality for scenario 2

Below are the results that evaluate the quality of the AD system reception. Figure 17 shows the normalized average scores for each of the questions described above.
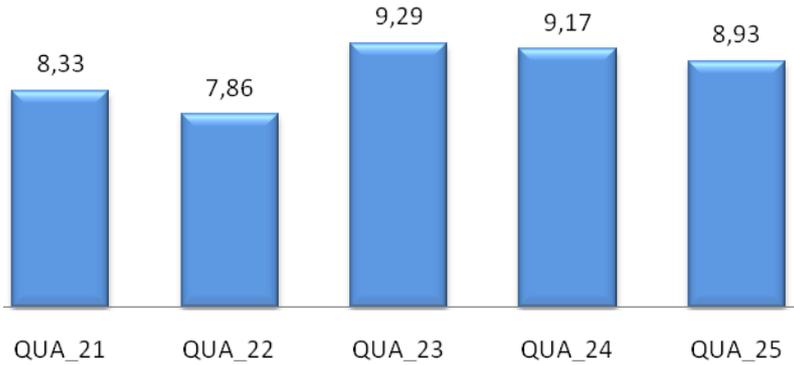
Figure 17: Normalized scores for the questions on the picture quality of scenario 2

In this scenario the participants have appreciated the opportunity to listen to the audio description of audiovisual content on an individual basis using a laptop while viewing the image on the television. The average overall score for the system quality is 8.68. With respect to the quality aspects evaluated, the synchronization between the video (and receiving apparatus viewed TV) and the audio (received on a laptop and listened to with headphones) was excellent (questions 21, 23 and 24). That is, the

participants have encountered no problems of synchronization. The speed of access to services has also been assessed very positively. The scores on this question (question 25 with 8.93 points) confirm this and no negative observations were obtained in this regard. Regarding question 22, it should be mentioned that, although there were some cuts and breaks in the audio, the rating remains positive. We conclude that these cuts are the result of the signal quality of the DTT antenna used during the tests. It seems that, not using a professional reception system (such as that found in neighbouring communities), the signal quality has been noticeably affected by noise and interference.

Finally, we asked the participants about the overall quality of the system using an absolute valuation. Specifically, question 26 asked:

QUA_26	In general, the system quality is: 5. Excellent 4. Good 3. Acceptable 2. Mediocre 1. Bad

<p align="center">Table 11: Question on an absolute scale for scenario 2</p>

The results are presented in Figure 18
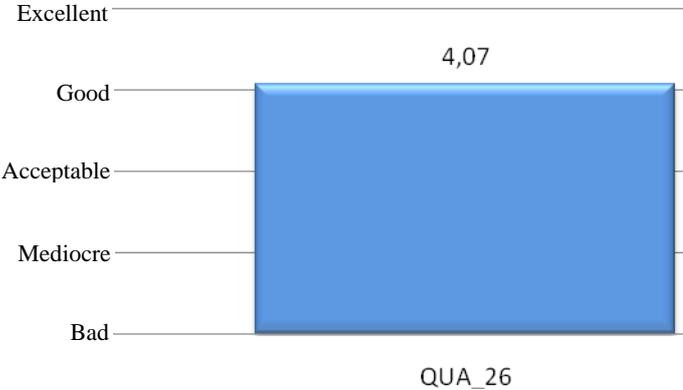
<p align="center">Figure 18: Normalized score for the question on quality of scenario 2</p>

The overall quality of the system was rated as "good" by most participants (64%) with an average score of 4.07 out of 5. This score can be considered good when we keep in mind that the scale was presented from 1 to 5. Regarding the distribution of responses from participants, Figure 19 shows that 22% of participants considered that the quality was "excellent", while the percentage of people who considered it only "acceptable" was 14%. This means that 86% of participants rated the quality of content presented in this scenario as "excellent" or "good." None of the participants rated the quality of the content as "mediocre" or "bad".
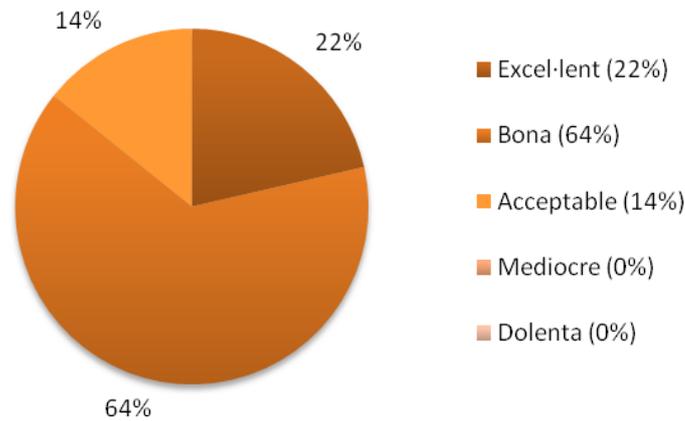
Figure 19: Frequency of responses to the question on the absolute quality of scenario 2

In short, a good evaluation for the AD system for reception validates this scenario as a possible solution for reconciling family circumstances where a person with visual impairment lives with unimpaired relatives.

c. Conclusions of the evaluation of scenario 2

In this scenario we have studied a solution for reconciling individual family reception of TV content with audio description. Though the population without visual impairments do not think is necessary and do not use audio description customisation of content to suit the user's needs and preferences seems very interesting according to the opinions collected. In this case, the evaluation has rated the content synchronized apparatus for TV and a laptop, where users can hear the audio description while watching TV with the rest of the family. The following summarizes the results of the evaluation of scenario 2 according to different scales.

Aspect rated	Score
Utility	**6,70**
Quality	**8,91**
Absolute quality (question 26 normalized and out of 10)	*8,14*

Table 12: Summary of scenario 2 evaluations

The main strength of the usefulness of this scenario is the possibility to receive audio description individually, while the rest of the family watches TV content without audio description. Although participants without visual impairments have not found it necessary to use audio description themselves (a result that was also collected for scenario 1) they have rated very highly this system as a possible solution to reconciling family watching and customisation.

Regarding the assessment of quality, and in particular the synchronization between the contents received on TV and the AD audio listened to with headphones attached to a laptop, participants

considered that it was excellent. There were no negative comments on this aspect and the average scores for quality and overall quality confirm this fact. In summary we can say that the quality of the system has been evaluated positively by participants.

Aspects taken into account include the need for the laptop with DTT receiver and headphones to be able to easily be integrated into the TV space and enjoy with family while avoiding an excess of cables or uncomfortable positions when using the laptop.

Finally, we collected the data and opinions suggesting that this scenario can be validated as a possible solution for reconciling family in cases where there is a person with visual impairment. Its acceptance has been demonstrated both in the technical aspects and in its application.

2.3. Scenario 3.1. On-demand AD: reception via a set-top box

The scenario of on-demand AD reception via a set-top box allows the user to select from a set menu of programs, and watch them on television over the Internet using a streaming IP channel.

a. Assessment of usability

This is done through questions about two aspects of usability, ease of use and interface. The ease of use was assessed through questions 1 to 7, while the interface was assessed through questions 8 to 12. The questions were as follows:

USA_01	Overall, I am satisfied with the usability of this system
USA_02	This system is simple to use
USA_03	I can easily complete the tasks that I have been requested to do
USA_04	I quickly complete tasks that I have been requested to do
USA_05	I can efficiently complete the tasks that I have been requested to do
USA_06	I feel comfortable using this system
USA_07	This system is easy to learn to use
USA_08	The information provided (text on the screen help text) is clear
USA_09	The organization of information in the system screens is clear
USA_10	Navigation through the menu screen is simple
USA_11	The information has been effective in helping me complete the tasks
USA_12	Interface of this system is pleasant

Table 12: Usability questions for scenario 3.1 with a numerical response scale

Figure 19 shows the results of each of the questions about usability. The values shown are mean normalized on a scale of 0 to 10, where 0 indicates "strongly disagree" and the value 10 indicates "totally agree". The colours refer to the subscale to which each question belongs.
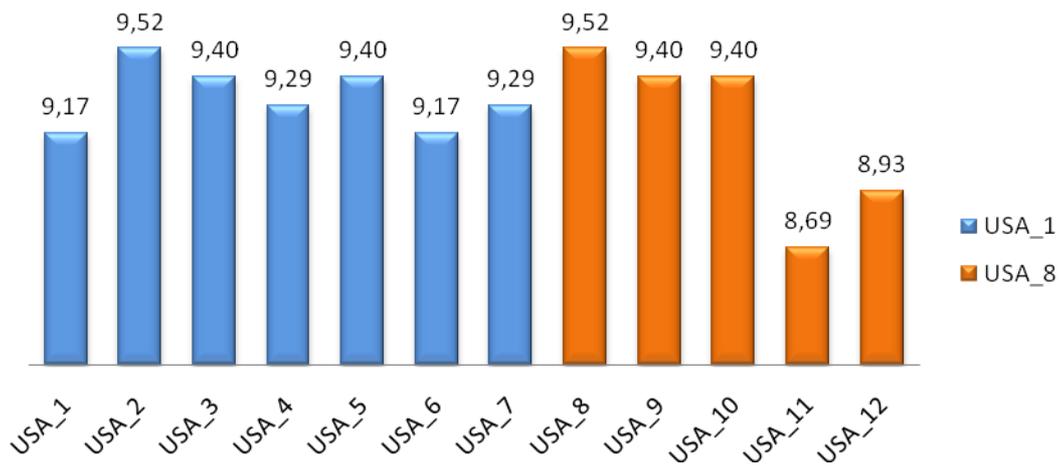
Figure 19: Normalized scores for the questions on the usability of scenario 3.1

The questionnaire data indicates that all aspects of the usability of the system on-demand AD reception via a set-top box get a good rating. As shown in Figure 19, all questions have an average score around 9 (minimum 8.69, maximum 9.52). The average usability score is 9.26 out of 10. Regarding the average values of the subscales, ease of use and the interface score 9.32 and 9.19 respectively. The general assessment is based on the scores for questions 11 and 12. In this sense, the data suggests that participants found the information presented and the system interface could not be improved, so we can say that the usability of the system is considered very good.

It is noteworthy that the variability of responses for each question has been minimal, indicating a high degree of consensus among participants.

b. Assessment of utility

The scale that evaluates the utility is made up of five Likert scale questions and 2 open questions. The Likert scale questions and open questions are given below.

UTI_13	This system has all the features and capabilities that I hoped it would have
UTI_14	In general, I think this service is useful
UTI_15	I have this service
UTI_16	If I had access to this service, I would use it
UTI_19	To what extent would your satisfaction increase if you had the service available on demand?

Table 13: Numerical scale questions on the use of scenario 3.1

UTI_17	What stands out as the main advantages of this service?
UTI_18	What stands out as the main disadvantages of this service?

Table 14: Numerical scale on the relative merits of scenario 3.1

Below we present the results of the questions for on-demand system AD reception via a set-top box. Figure 20 shows the normalized average scores for each question on utility with a numerical scale. As always, the value 0 indicates "strongly disagree" and the value 10 indicates "totally agree".
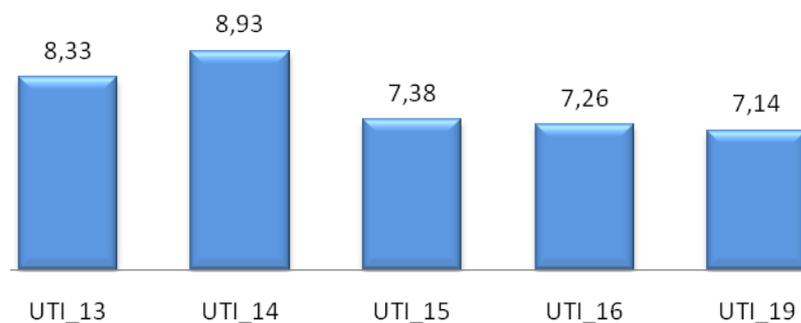

Figure 20: Normalized scores for the questions on the use of scenario 3.1

The average overall score for the usefulness of on-demand system AD reception via a set-top box is 7.8. Thus, participants have positively assessed the usefulness of the service. As in previous systems evaluated, the question on the general service utility (question 14) is the best valued and has a higher degree of agreement among participants. Then, with the highest scores, there is the question about the system functions (question 13) with 8.33 points. Questions 15 and 16 regarding whether the participants would like to have the service and if they use it, also receive a positive assessment. Finally, question 19, which refers to the degree of satisfaction of having the service, the data shows a tendency to increase the satisfaction of participants, although not very marked.

In the open-answer questions designed to assess the pros and cons of the system on-demand reception in set-top box, we have obtained the following answers. The advantages of using the system can be grouped as follows. On one hand, participants repeatedly highlighted the flexibility of time that allows the availability at any time previously issued chapters, allowing viewers to adapt to the time schedule. However, it also highlights the fact that access is instant (streaming), so there is no wait time to begin to see the contents. In this sense, a reply stating "access at any time without waiting for the programming required" or "not necessary to download the chapter and is fast."

Meanwhile, the feedback on the negative aspects of the system highlights the fact that the display does not have access to a time bar to fast forward or rewind the content (using the remote control, for example). Another point to mention is that, in the application, it is necessary to have to remember the chapter number, since it is the only reference users have to content. Regarding this point, observations were made related to the information provided about the contents. Some participants positively assessed the fact that it offered more information about the chapters (for example, a brief written summary).

c. Assessment of the quality

To evaluate the subjective quality of this scenario, the participants were asked six Likert scale questions and a question with an absolute scale of 5 levels. The Likert scale questions are shown below:

QUA_20	The image quality is good.
QUA_21	The audio quality is good.
QUA_22	There are no cuts or breaks.
QUA_23	After selecting the service, access to content is quick.
QUA_24	The quality is the same as in the system of DTT.
QUA_26	Overall, I am satisfied with this system.

Table 15: Numerical scale questions on scenario 3.1

Below are the results that evaluate the quality of on-demand system AD reception via a set-top box. Figure 21 shows the normalized average scores for each of the questions described above.
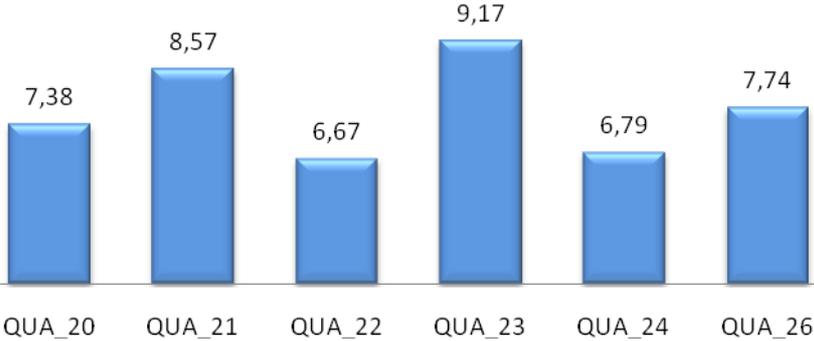
Figure 21: Normalized scores for the questions on the quality of scenario 3.1

The global average score for on-demand AD reception via a set-top box is 7.72. Note that the satisfaction with the above scenario is still remarkable, indicating a clear satisfaction with the system.

With respect to the elements better rated, speed of service access has been rated very positively, being the determining factor in assessing the quality of this scenario. In this scenario, as has happened in the first scenario, the score for the image quality has been affected by the cuts and breaks that have occurred during the display of content (question 22, with 6.67 points). In the same way as the content of Live Streaming Internet TV, videos accessed by request and sent streaming also occasionally had some cuts and jumps. These have resulted directly in the assessment of image quality (question 20, with 7.38 points) and of the comparative quality with respect to DTT (question 24, with 6.79 points). It should also be noted that the assessment of audio quality was not affected by these cuts.

In question 25 the participants were asked about the general quality of the system of AD on-demand and its reception via a set-top box.

	In general, the system quality is:
QUA_25	5. Excellent 4. Good 3. Acceptable 2. Mediocre 1. Bad

Table 15: Question on an absolute scale on scenario 3.1

The results are shown below.
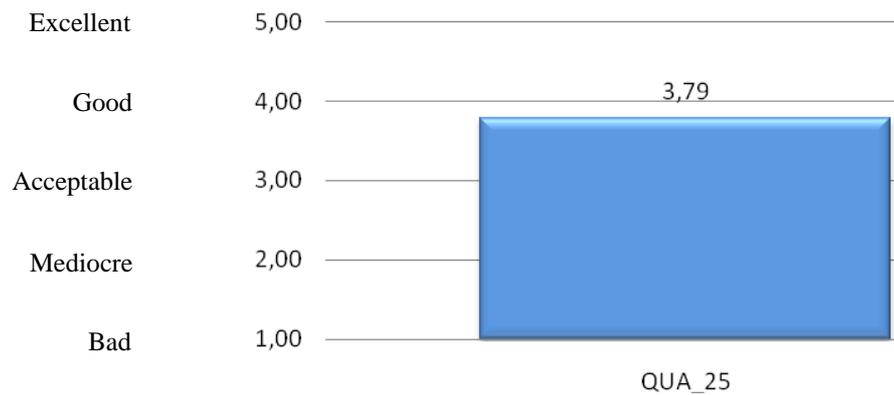
Figure 22: Score for the absolute standard question on the quality of scenario 3.1
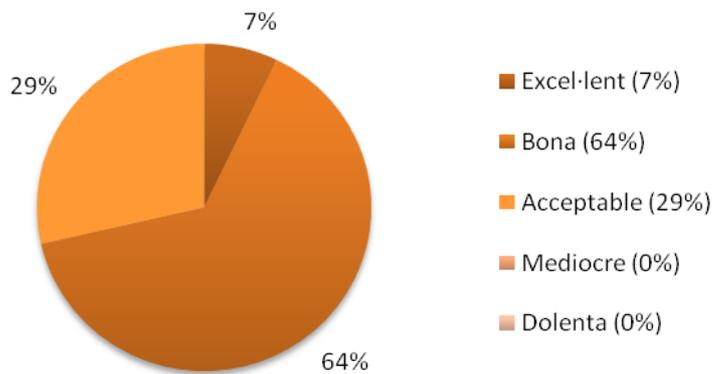
Figure 23: Frequency of responses to the absolute scale question on the quality of scenario 3.1

The average score from assessing the overall quality of this system was 3.79, which represents a quality rating of "good" to "acceptable." In this case more participants have found the quality "acceptable" (29%) than those that have found it to be "excellent" (7%), thus reducing the overall average.

d. Conclusions of the evaluation of scenario 3.1

This scenario of a transmission system for on-demand Internet TV content, also called on-demand AD reception via a set-top box was evaluated subjectively. In the case of this scenario the contents have been selected and reproduced from a set-top box with Internet access and connected to the TV. As in other scenarios, an important part was to assess the presence of audio description in the contents. The following summarizes the results of the evaluation stage according to different criteria.

Aspect rated	Score
Usability	**9,26**
Ease of use	*9,32*
Graphic interface	*9,19*
Utility	**7,80**
Quality	**7,72**
Quality (Question 25 average score out of 10 and normalized)	*7,58*

Table 17: Summary of the evaluation of scenario 3.1

Overall, the assessment of this scenario has been favourable. Similarly to the results obtained in the first scenario, we stress that usability has been the most valued facet, both in terms of ease of use and the graphical user interface. Participants emphasized the ease and speed of access to on-demand video service with AD. It is important to remember that the graphical interface used to access the service is the same as that of the first scenario. In this sense, it is not surprising that the ratings have been similar in this respect between the two scenarios.

The evaluation of the usefulness of the scenario has been influenced by subjective considerations of the participants. A service with audio description may in some cases be used by users who do not need AD and do not have someone at home that needs it. On the other hand, the possibility of acquiring over the Internet contents anytime and anywhere which have been previously been released was rated very useful by the participants. These aspects make this scenario a very attractive solution in today's information society. This was crucial in increasing the score for the usefulness of this scenario.

Regarding quality, the participants have found the same effects as the first scenario, a fact expected if one takes into account that the transmission system is almost the same in both scenarios (and therefore both systems are affected by the same problems). On the other hand, participants positively rated the speed of access to AD.

As has happened with the first scenario, taking into account the potential improvements in service quality by avoiding unwanted cuts and breaks, the system demonstrated has had its feasibility endorsed by the test participants who have indicated an acceptance of it.

2.4. Scenario 3.2. On demand AD: Consumer PC

For the scenario of On-demand AP via a Consumer PC participants were asked to download the files containing the content onto a PC and view them on the same computer. For playback the content has been reproduced on a laptop with VLC playback software.

a. Assessment of usability

As in other scenarios, evaluation of the usability is based on questions about the ease of use and the interface. The evaluation of ease of use is through questions 1 to 7, while the evaluation of the interface is through questions 8 to 12. The questions are presented in the Table 18 below.

USA_01	Overall, I am satisfied with the usability of this system
USA_02	This system is simple to use
USA_03	I can easily complete the tasks that I have been requested to do
USA_04	I quickly complete the tasks that I have been requested to do
USA_05	I can efficiently complete the tasks that I have been requested to do
USA_06	I feel comfortable using this system
USA_07	This system is easy to learn to use
USA_08	The information provided (text on the screen help text) is clear
USA_09	The organization of information in the system screens is clear
USA_10	Navigation through the menu screen is simple
USA_11	The information has been effective in helping me complete the tasks
USA_12	The interface of this system is pleasant

Table 18: Numerical scale questions about the usability of scenario 3.2

The results for each numerical scale question on usability are presented below. Values presented are mean normalized, the value 0 meaning "strongly disagree" and the value 10 "totally agree". The colours refer to the subscales.
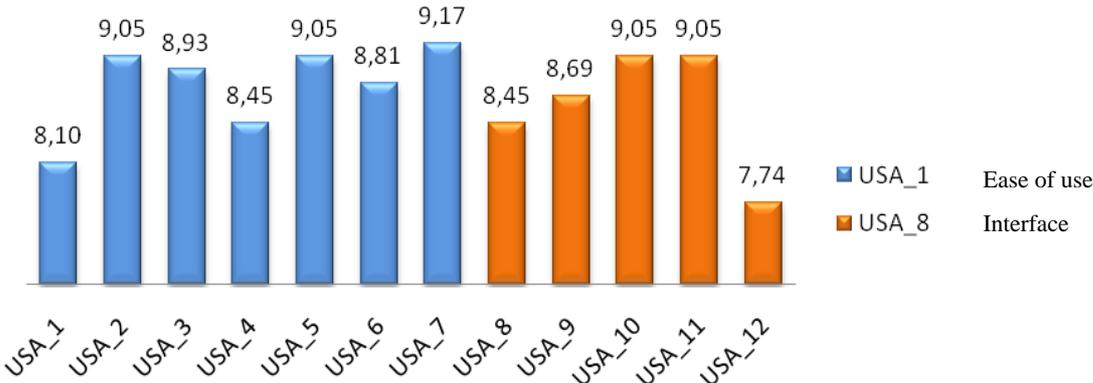

Figure 24: Normalized scores for the questions on the usability of scenario 3.2

The mean values indicate very positive aspects that make the on-demand AP: Consumer PC system useful. The overall average score for usability is 8.71 out of 10 (the minimum scored is 7.74, the maximum 9.17). If you look at the subscales, the subscale ease of use gets an average score of 8.79, while the subscale interface gets an average score of 8.5. Thus, the evaluations regarding the usability of the system can be described as very positive. If we look in more detail at the scores obtained, we observe the existence of a question with a score slightly below the rest. This question (question 12) refers to whether the system interface is nice. Its rating is 7.74, so it seems that user satisfaction is a bit smaller than with the rest of the aspects evaluated. With respect to the validity of the data, it is important to note that the averages presented are often obtained from data distributions with little variability, as happened in previous scenarios.

b. Assessment of utility

The evaluation of the usefulness of the system is based on asking the participants four Likert scale questions and 2 open questions. The Likert scale questions and the open question are given below.

UTI_13	This system has all the features and capabilities that I hoped for
UTI_14	In general, I think this service is useful
UTI_15	I have this service
UTI_16	If I had access to this service I would use it

Table 19: Numerical scale questions on the use of scenario 3.2

UTI_17	What stands out as the main advantages of this service?
UTI_18	What stands out as the main disadvantages of this service?

Table 20: Numerical scale questions on the relative merits of scenario 3.2

Figure 25 shows the results that evaluate the usefulness of the service on-demand AP: Consumer PC. The figure shows the normalized average scores for each of the numerical scale questions above.
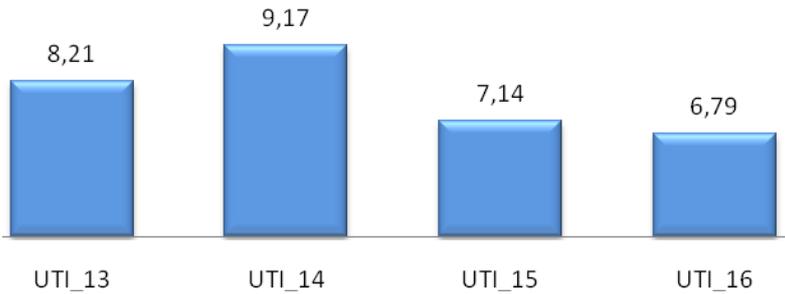
Figure 25: Normalized scores for the questions on the utility of scenario 3.2

The average overall score for the usefulness of the system on-demand AP: Consumer PC is 7.82. This data indicates that the overall utility is positive.

The question regarding the overall utility (question 14) and the question on the functionality of the system (question 13) scored, 9.17 and 8.21 points respectively. In the same way as in the evaluation of

previous systems, questions 15 and 16 regarding whether the participants would like to have the service and if they use it, get lower scores, with 7.14 and 6.79 points, respectively. That is, the data seem to point to the system being assessed as very useful with functions that are good to have, and that the "need" to have it or use it are valued positively, although to a lesser extent, Figures 26 and 27 show the distribution of answers to these questions.

Figure 26: Frequency of responses to the question "I have this service"
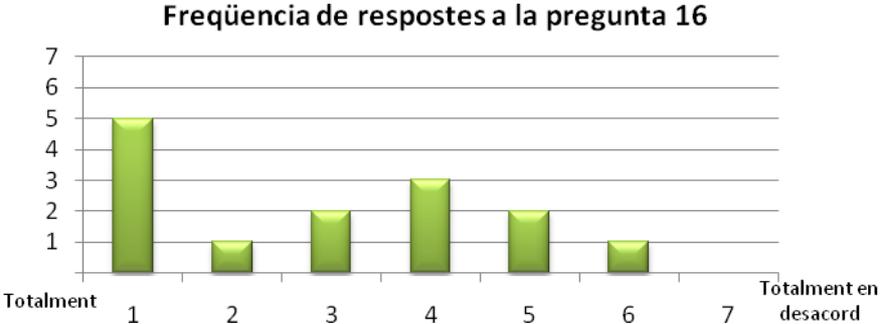
Figure 27: Frequency of responses to the question "If I had access to this service, I would use it."

Participants highlighted a number of advantages of the system in their responses to the open questions. The most important point made is the flexible hours of the service on-demand AP: Consumer PC. The flexibility of location is highlighted, namely, the ability to download the file at a given time for display on a portable device (for example, downloading from a desktop computer and viewing it on a smartphone). Another positive point to note is mentioned which is a guarantee of good quality download files with this system. Moreover, the main negative aspect that must be mentioned is the waiting time for downloading. This aspect is the most negatively rated, taking into account that domestic connections are generally slower.

c. Quality

To evaluate the subjective quality of this scenario, the volunteers were asked six Likert scale questions and a question with answers on an absolute scale of 5 levels. The numerical scale questions are shown in the table below.

QUA_20	The image quality is good.
QUA_21	The audio quality is good.
QUA_22	There are no cuts or breaks.
QUA_23	After selecting the service, access to content is quick.
QUA_24	The quality is the same as in the system of DTT
QUA_26	Overall, I am satisfied with this system

Table 21: Numerical scale questions on scenario 3.2

Below are the results that evaluate the quality. Figure 28 shows the normalized average scores for each of the questions above.
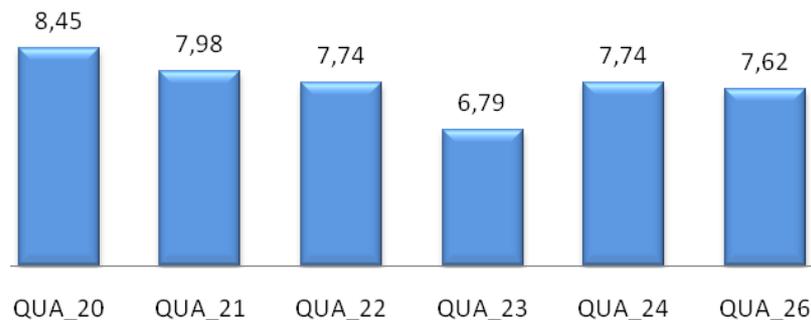

Figure 28: Normalized scores for the questions on the quality of scenario 3.2

The overall average quality score of this scenario was 7.72 points. This value suggests that the participants judged that the quality provided by the system is good.

The most penalized aspect is the speed of access to content. This is mainly due to the downloading time of the video file after being selected by the user: 5 minutes to download a 630MB episode of the "La Riera". This waiting time can be considered reasonable for any system video download over the Internet with existing network infrastructures. However, it must be remembered that in the experimental stage the download time is "dead" time-out, while in a real scenario a user may fill the waiting time performing other Internet activities. In this sense, it reflects well the question about the speed of access (question 23, with a score of 6.79), which is the lowest score for this aspect in the different scenarios. This is what has affected mainly the satisfaction of users with the system (question 26, with a score of 7.62). Some of the comments made by participants in this regard have been "We have to wait for the download to see it," The time it takes to de-load can be long or "Time to download content can sometimes be high."

The image quality and the audio has not been affected by the existence of cuts or jumps due to the Internet or playing on a set-top box because the contents are reproduced locally on a laptop.

In the question on quality participants were asked about the general quality of the on-demand system AP: Consumer PC. Specifically, the 25 questions were asked:

QUA_25	In general, the quality system is: 　　5. Excellent 　　4. Good 　　3. Acceptable 　　2. Mediocre 　　1. Bad

Table 22: Question on an absolute scale on the quality of scenario 3.2

Figure 29 represents the results question 25 on a scale of 5 points.
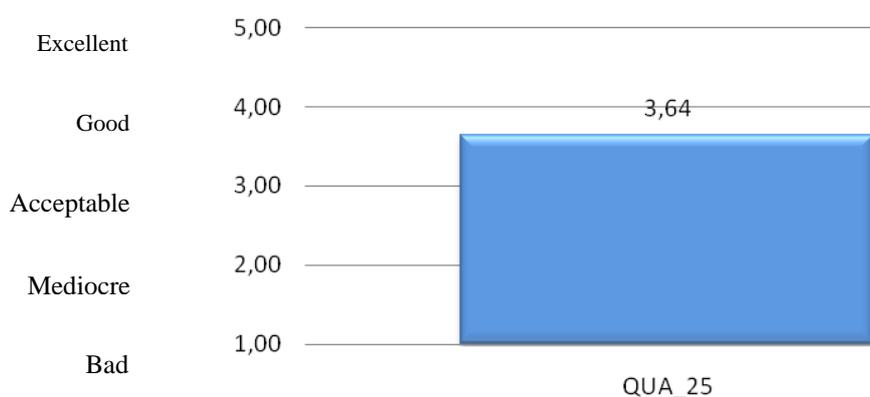
Figure 29: Score for the absolute standard question on the quality of scenario 3.2
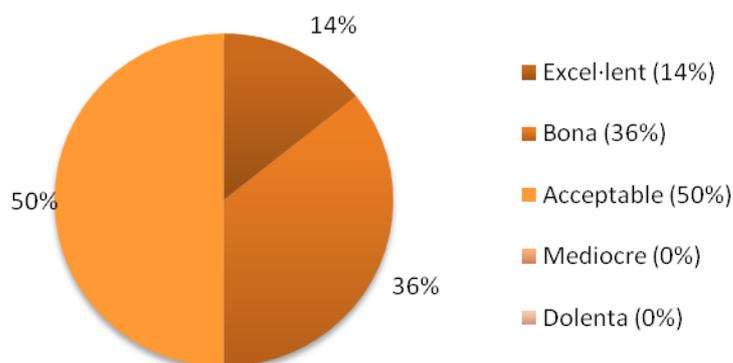
Figure 30: Frequency of responses to the absolute scale question on the quality of scenario 3.2

The download time of content has been the main factor that has contributed to the overall quality rating. Half of the participants have considered the quality of the system "acceptable" while the rest considered the quality "Good" or "Excellent." Given that the download time was practically the same for each participant, these differences could be explained by the consideration of participant's subjective view of the time "dedicated" to this activity. The 5 minutes of waiting time can be excessive if not complemented by another activity or alternative services.

d. Conclusions of the evaluation of scenario 3.2

The scores obtained in the evaluation of scenario 3.2 according to different scales of study are presented in Table 23 below.

Aspect rated	Score
Usability	**8,71**
Ease of use	*8,79*
Graphic interface	*8,50*
Utility	**7,82**
Quality	**7,72**
Quality (Question 25 normalized out of 10)	*7,28*

Table 23: Summary of the ratings of scenario 3.2

Data from ratings obtained from the different aspects tell us that the participants have rated the service positively. If you look at the different aspects that have been evaluated, we can observe how the usability of the system is highly appreciated by the participants, with a score closest to 9. Meanwhile, both the utility and the quality scores are close to 8.

The usability score indicates that the realization of the tasks requested in this scenario has been performed satisfactorily, both with regards to ease of use and handling of the interface. The data and comments collected to evaluate the interface indicate that participants believe that the interface used may be improved. At the same time, participants have suggested that the introduction of audio description of the menus of the application would increase the usability of the system and make it more accessible.

Regarding utility, the data show that the usefulness of this system has been judged positively. Being able to download the video file with audio description is an improvement in view of the flexible hours of consuming the AD it offers while increasing the alternative content reception by the viewer. However, being able to reproduce the contents in other portable devices has been rated as very useful. In relation to reception on other devices, mention has been made of the possibility of providing the content for playback in various formats adapted to different mobile platforms (smartphones, i-pad, etc.) thus optimizing the use of the resources of these platforms.

Concerning the quality of the system, the results indicate that participants appreciate the level of the quality of the system. The display of the contents was made without incident, and evaluations by the participants regarding the level of video quality and sound was good. In this regard, participants stressed that the guarantee of picture quality and sound represents the availability of this service directly from the producer compared to obtaining content from third parties. However, the waiting for the download of contents has resulted in a lower overall score for the utility of the system. As

explained in the section devoted to assessing the quality, the download time of the tests was around 100Mb. If we take into account domestic connections that currently have a much lower speed, it would be advisable to consider this factor when making future deployment decisions. In this regard, the implementation of a more efficient download manager that allows multiple files to be downloaded at once would be a helpful tool if you want to download more than one chapter (for example, download the chapters for the week as a whole). This tool could be a substantial change to the user experience, helping to better organize download times.

2.5.   Scenario 3.3. On demand AD: reproduction via a *Podcast*

The scenario of on-demand AD podcasting allows users to download audio files on a PC and listen to them on the same computer or a portable media device. To reproduce the contents in the test a laptop with VLC software was used.

a. Assessment of usability

As in previous scenarios, the questionnaire is composed of questions about usability, ease of use and interface. The numerical scale questions asked are presented in Table 24.

USA_01	Overall, I am satisfied with the usability of this system
USA_02	This system is simple to use
USA_03	I can easily complete the tasks that I have been requested to do
USA_04	I quickly complete tasks that I have been requested to do
USA_05	I can efficiently complete the tasks that I have been requested to do
USA_06	I feel comfortable using this system
USA_07	This system is easy to learn to use
USA_08	The information provided (text on the screen help text) is clear
USA_09	The organization of information in the system screens is clear
USA_10	Navigation through the menu screen is simple
USA_11	The information has been effective in helping me complete the tasks
USA_12	The interface of this system is pleasant

Table 24: Numerical scale questions about the usability of scenario 3.3

Figure 31 below shows the results for each numerical scale question on the usability of scenario 3.2.
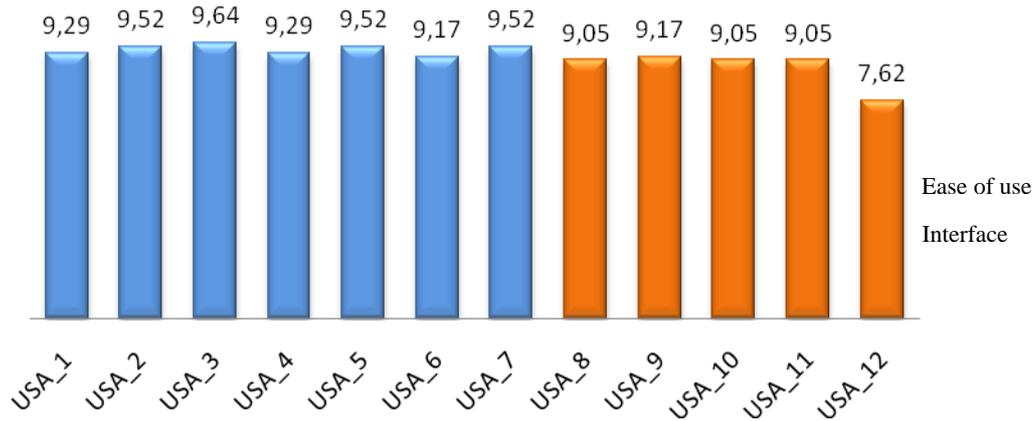
Figure 31: Normalized scores for the numerical scale questions on the usability of scenario 3.3

The results of the evaluation of the utility system On-Demand AD podcasts indicates that the participants very positively evaluated all aspects of usability. All scores (except one) are above 9 points, indicating an excellent level of usability. The average global score is 9.15 points. The mean of the subscales are 9.42 for the scale of ease of use and 8.78 for the interface. The scores given to each question were very similar among the participants. It is necessary to mention the score for question 12 about whether the interface is nice. The value of 7.62 suggests that the aesthetics of the interface could be improved. For example, some evaluation participants rated the Pleasantness of the interface with a "neither agree nor disagree".

b. Assessment of the utility

To assess to what degree the system on-demand AD podcasts is helpful to the participants, we asked five Likert scale questions and 3 open questions. The questions are shown in the following tables.

UTI_13	This system has all the features and capabilities that I hoped it would have
UTI_14	In general, I think this service is useful
UTI_15	I have this service
UTI_16	If I had access to this service, I would use it

Table 25: Numerical scale questions about the utility of scenario 3.3

UTI_17	What stand out as the main advantages of this service?
UTI_18	What stand out as the main disadvantages of this service?

Table 26: Absolute scale questions about the utility of scenario 3.3

Figure 32 below shows the normalized average scores for each of the numerical scale questions described above.
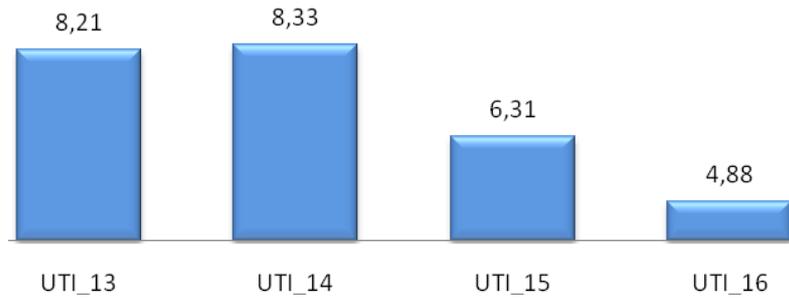
Figure 32: Normalized scores for the questions on the utility of scenario 3.3

The average score for the utility of the scenario on-demand AD podcasts is 6.93. If we review the data obtained for each question, we observed that questions 13 and 14 relating to the functionality and utility score higher than 8. Moreover, questions 15 and 16 are rated lower, due to a more divided response indicating that, in general, participants in the test showed a moderate interest in having this service and in its use (as has been observed in previous scenarios).

In the open questions on the positive and negative aspects of this system, the following observations have been collected: Regarding the advantages of the system on-demand AP: podcasts, time flexibility of use has been highlighted. For example, one volunteer said "people following the series can hear it in the car or doing other things," another comment was "It is very useful for blind people. This podcast may be enjoyed on the train." This scenario also allows audio reproduction for sighted persons while there are activities that require their visual attention such as driving or playing sports.

Regarding the negative aspects, the waiting time involved in downloading was mentioned, although this time is probably small given the relatively small size of audio files. A suggestion to improve the system was "it would be an improvement to include an audio menu system for easy access for the blind."

c. Assessment of Quality

To evaluate the subjective quality of this scenario six Likert scale questions were asked and a question with an absolute scale of 5 levels. The numerical scale questions contained in the questionnaire are given in Table 27 below.

QUA_21	The audio quality is good.
QUA_22	There are no cuts or breaks.
QUA_23	After selecting the service, access to content is quick.
QUA_26	Overall, I am satisfied with this system.

Table 27: Numerical scale questions on the quality of scenario 3.3

Figure 33 shows the normalized average scores for each of the questions described above.
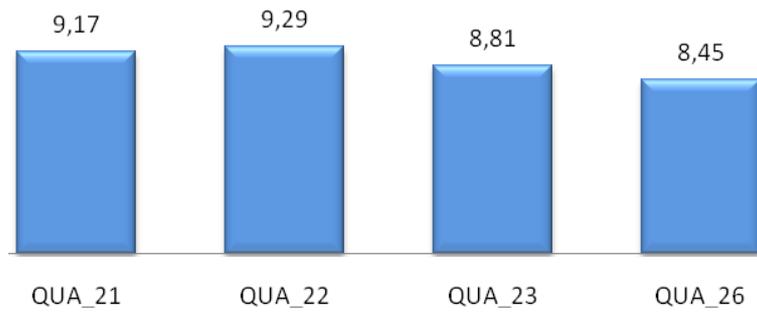
Figure 33: Normalized average scores for the quality of scenario 3.3

For this scenario, the assessment of quality has clearly been positive both in regard to audio quality (e.g. possible jumps and cuts) and speed of access to content. The overall average utility score is 8.93 points. If we analyze the numerical scale questions, we can observe that the participants have been very satisfied with the quality of the system, standing out as more significant are the absence of cuts and breaks (question 22) and the quality of the audio (question 21). Also with a very good score is the participants' assessment of the time taken to access content (question 23). Participants the waiting time for downloading very acceptable, which was about 15 seconds for an audio file of about 40MB. The question on the satisfaction of the participant with the system (question 26) yielded a score of 8.45.

Below the results of question 25 which asked the participants to give their assessment of the overall quality of the system are given.

QUA_25	In general, the quality of the system is
	  5. Excellent
	  4. Good
	  3. Acceptable
	  2. Mediocre
	   1. Bad

Table 28: Question on an absolute scale on the quality of scenario 3.3
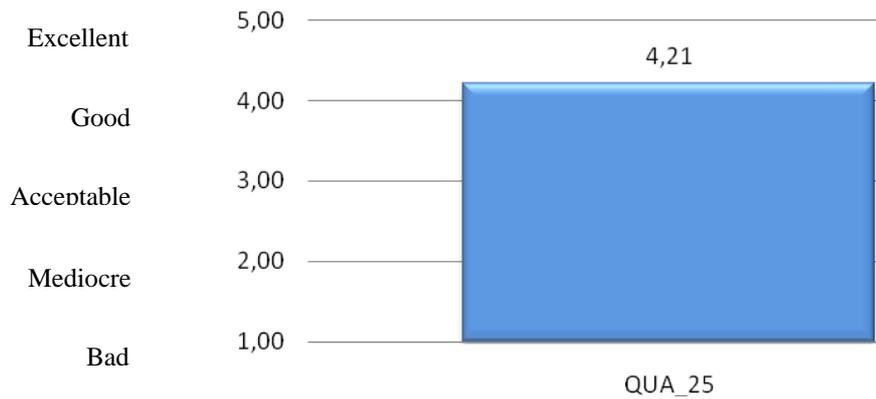
Figure 34: Score for the absolute standard question on the quality of scenario 3.3
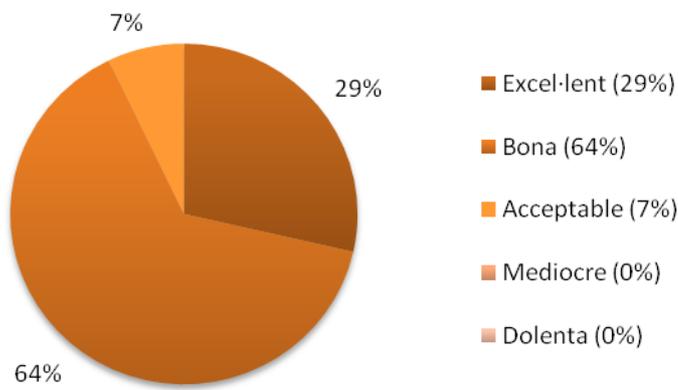
Figure 35: Frequency of responses to the absolute scale question on the quality of scenario 3.3

In evaluating the subjective quality of this scenario, where participants download and listen to an audio podcast of a chapter of "La Riera", 93% of participants considered that the content quality was" good " or "excellent" while the remaining 7% considered it "acceptable." Specifically, it has been rated as "excellent" by about a quarter of participants (29%) and "good" by most respondents (64%). As in previous scenarios, no participant considered that the quality of content displayed was "mediocre" or "bad."

As regards general observations, these were concerning more with the audio description script than with the system evaluated, it should be noted that some participants have commented on the difficulty of following the plot in podcast. For example, it was mentioned that the scene changes are sometimes difficult to follow, as one participant commented "I am satisfied with the system, except when there is a change of scene." Also mentioned was the fact that, the chapter "La Riera" began with a summary of the previous chapter, despite some of the participants not really knowing the previous chapter. By way of observation, one participant commented "It is not clear when the audio describer starts talking about the chapter itself". In this sense, we thought it interesting that the comments focused on the audio files used in the evaluation of this scenario.

d. Conclusions of the evaluation of scenario 3.3

The scores obtained in the evaluation of scenario 3.3 according to different scales of study are presented in Table 29 below.

Aspect rated	Score
Usability	**9,15**
Ease of use	*9,42*
Graphic interface	*8,78*
Utility	**6,93**
Quality	**8,93**
Quality (Question 25 normalized out of 10)	*8,42*

Table 29: Summary of the scenario 3.3 evaluations

The analysis of the results indicates a very positive evaluation by the participants. Paying special attention to the scores, the data shows how the usability has been rated the best aspect of this scenario. Specifically, the ease of use subscale has a score of 9.42. Also highly valued was the interface with 8.78 points. These scores indicate that our system is easy to use when doing a podcast download. Regarding the assessment of usability it should be noted that, given the fact that the application used is the same as that of scenario 3.2 (On-demand AP: consumer PC), the ratings obtained are very similar. In this sense, the assessment regarding the interface has been generally very positive, with values always higher than 8.5. As for the previous scenario, the score provided for the Pleasantness of the interface suggests that its aesthetics can be improved. As before, it has been identified that the introduction of audio description of the application menus will facilitate accessibility.

The flexible hours for AD reception that this system allows have been highlighted. However, the analysis suggests that the utility of download podcasts is less than downloading the episodes on video, noting the scale of the questions, it can be seen that the questions about whether the participants would like to have the system and if they use it are rated with a lower score.

In assessing the quality of the system the participants rated very highly both the audio quality of the broadcast and the speed of access to the contents. You can also see the participant satisfaction in their responses to the question about the overall quality.

As a final remark, as was pointed out above, it must be remembered that most of the participants echoed the difficulty in following the chapters in the podcast category. We include this interesting discussion because audio description chapters can also be an option among those without visual impairments who want to follow a series while doing other activities (driving, travelling on public transport, sports, etc.). A possible explanation for these problems in understanding may be due to the some people not regularly watching such content. However, remember that TV content converted to

podcasts shows some substantial differences in the structure and content of contents that were designed exclusively for podcast or the radio, and that lack of understanding of the visual content can be a disadvantage.

Conclusions

The results indicate that the subjective assessment of the viability of many services indicates that they are feasible. The numerical data, together with the views collected, provide information on different aspects to consider when planning the deployment of these services.

By way of summary, Table 30 represents the aspects that were highlighted as positive aspects as well as the subject of improvement for each service. In the table the green dots indicate that participants have positively assessed the service in a particular aspect. The orange dots refer to issues that, while acceptable, are rated as needing improvement in some sense. Finally, red dots indicate those aspects which can be critical when considering the deployment of the service.

Aspect rated	Scenario 1	Scenario 2	Scenario 3.1	Scenario 3.2	Scenario 3.3
Usability	🟢	▬	🟢	🟢	🟢
Ease of use	🟢	▬	🟢	🟢	🟢
Graphic interface	🟠	▬	🟠	🟠	🟠
Utility	🟢	🟠	🟢	🟢	🟠
Quality	🔴	🟢	🔴	🟠	🟢

Table 30: Summary highlights of the factors in each scenario, where a green dot indicates that the variable has been featured positively in a given scenario, an orange dot indicates that there are aspects we can improve on not considered as crucial in the deployment of technology and a red dot indicates the presence of crucial aspects to improve.

Regarding the data in the table we can see some common patterns with regard to the issues rated (i.e., usability, usefulness, and quality). In this sense, the usability (and specifically, the ease of use) received very positive reviews in a systematic way. It is also interesting to note how the ratings on the GUI are rated as subjects for improvement, especially by including audio descriptions of sound menus. The ratings were most critical for systems involving the use of streaming. Although it must be remembered that the services evaluated are not comparable with each other, we can see that patterns emerge in the data that have a real foundation in the sense that the ratings are consistent in what they assess (e.g. ease of use). However, the fact that the participants have expressed similar opinions systematically gives validity to the results.

An overview and summary of the aspects evaluated in the various scenarios is provided in Table 30.

Regarding the evaluation of scenario 1, Live Streaming Internet TV, we can say that the usability and usefulness of the system are good, while we should review the final quality of the reception of content, which is in part affected by the presence of cuts and breaks in the image. However, the usefulness of having the same service for live Internet broadcast by Freeview has been positively highlighted.

Regarding scenario 2, AD use, both the utility and quality are satisfactory overall. As the participants without visual impairments have not been shown to be very likely to watch content with AD, the system of personalized TV reception for people with visual difficulties (on a portable device) was rated very useful. This personalization and improvement in family bonding that this solution offers have been identified as the most positive aspects of the scenario by the participants. The suggestion to consider the possibility of using a smaller device for receiving DTT with audio description to take into account the convenience of the user is useful.

Scenario 3.1, on-demand AD reception via a set-top box is well evaluated in terms of usability and usefulness. However, we should take into account aspects of quality of service in the reception of streaming content to try to reduce the appearance of cuts and match the quality of reception with the DTT signal. In any event, the participants also emphasized the usefulness of being able to access content at any time.

The assessment of Scenario 3.2, On-demand AD: consumer PC has been very positively received both in usability and utility, which has highlighted the possibility of accessing content anytime and anywhere. The assessment of quality has been affected by the download time of content (variable depending on the bandwidth of the Internet access), but not the quality of content downloaded. As a suggestion, some participants discussed the possibility of adapting this content for mobile devices.

For scenario 3.3, On-demand AD podcast reception data indicates that the utility is valued less than for some earlier scenarios, mainly because the participants found it difficult to continue the series with only the audio (podcast), even with the help of audio description. In addition, both the usability and quality have been highly valued.

In summary, the emerging services analyzed in this work are technically feasible but must take into account some aspects of service quality to enhance the user experience for consumers. The usability of these services has been clearly validated with respect to ease of use and suggestions for improvements regarding the GUI. The utility has also been validated by the participants of audio description which has been rated very highly, not only as a tool to make content more accessible, but also as a means of bringing family together. With this data we can conclude that the proposed services are widely accepted by consumers, are easily accessible and have sufficient quality to merit their deployment.

References

- ITU-R Recommendation BT.500-11 (2002): "Methodology for the subjective assessment of the quality of television pictures." International Telecommunication Union, Geneva, Switzerland.

- ITU-T Recommendation P.910 (1996): "Subjective video quality assessment methods for multimedia applications." International Telecommunication Union, Geneva, Switzerland.

- Lewis, James R. (1995) 'IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use', International Journal of Human-Computer Interaction, 7: 1, 57 — 78.

Annexes

Incidents

During the execution of the tests several incidents that have affected the normal course of testing were recorded.

Technical Issues

After the test set up was in order the teams carrying out the scenarios worked well during testing with few notable exceptions. Among these we must mention those relating to the set-top box, which in all cases could be solved by restarting the device.

One of the specific incidents on the set-top box was we believe due to the connection between this device and the video server streaming TVC. Although the interactive application was connecting correctly, the set-top box could not even access or display the content or live content on demand. After restarting the set-top box several times (disconnecting it) we contacted TVC to verify that the service was not a problem, so we confirmed there was a problem. Without having found the source of the error, the set-top box was left disconnected for about five minutes, and this solved the incident.

The remaining technical issues have been related to receiving the signal, for Freeview and the Internet, breaks and cuts occasionally detracted from the enjoyment of the contents. In DTT coverage the antennas used at times caused similar problems. This forced a suspension of a portion of one of the sessions of the scenario concerned. This session was complemented later with another subject. In Internet services, however, we cannot be sure of the origin of these small cuts and breaks, but we think that may have been caused by network congestion. In any case, the necessary explanations were given to the subjects to take into account these incidents and their origin.

Incidents from logistics

As an important part of the contents used in the evaluation were broadcast live on TVC when scheduled (15:50 am) until a certain date (16 July), the incidents have mainly been a result of logistical and timing constraints.

Program schedules dictated the planning and timing of the experiments. To adjust ourselves to the timing experiments were held for subjects with sufficient time to give the necessary instructions before the program started. However, the delay in the arrival of some participants required an adjustment in the timing of these instructions.

Finally, as is customary in such experiments, the management of the subjects of was crucial for the proper performance of the tests. Despite having had an accurate diary of people and dates, subjects' absence at the last minute required replacements to found quickly. It should be kept in mind that the subjects' participation in the experiment s was entirely voluntary.

Quetionnaire

Durant el mes de juliol tenen lloc a la UAB unes proves per tal de saber

què en pensa la gent sobre un conjunt de nous serveis de televisió que TV3 vol oferir properament.

Aquests nous serveis tenen l'objectiu de **facilitar que les persones cegues accedeixin als continguts audiovisuals**. Els serveis oferiran diverses alternatives per a poder accedir a tots aquells programes, sèries i pel·lícules que tinguin una Audiodescripció per a persones cegues. L'Audiodescripció permet escoltar una descripció que explica tot allò que es veu en l'escena.

Per a **valorar si aquest serveis son prou útils i funcionen correctament**, ens cal la col·laboració de persones com tu!

Les proves consisteixen en veure alguns fragments de sèries que s'emeten per TV3 i valorar-ne aspectes com la qualitat de l'emissió o la facilitat per accedir al servei d'audiodescripció.

Després de veure els fragments, hauràs de valorar diferents aspectes del servei mitjançant qüestionaris molt senzills de respondre.

La teva col·laboració és molt important!

....

A les 15.25h del dia acordat t'esperaran a l'entrada de l'edifici **Mòdul de Recerca A** (davant la Biblioteca d'Humanitats).

Les proves duren al voltant d'una hora.

Es fan individualment (tot i que hi pot haver algun altre participant a la mateixa sala).

No es faran preguntes sobre detalls dels continguts visualitzats, ni preguntes complicades o personals.

Moltes gràcies!

....

Consentiment escrit

INFORMACIO DE L'EXPERIMENT

L'objectiu de les proves que es portaran a terme a la Universitat Autònoma de Barcelona (UAB) és l'avaluació subjectiva de la usabilitat i utilitat de diferents serveis emergents per audiodescripció. Els serveis a avaluar es concreten en tres escenaris diferents proposats per Televisió de Catalunya (TV3). Aquests escenaris ofereixen a l'espectador serveis diferents, dels quals els participants en els tests n'avaluaran la usabilitat i utilitat.

Els objectius de les proves d'avaluació impliquen donar resposta quantitativa i/o qualitativa a les següents qüestions generals:

- En quin grau és interessant per als espectadors el nou servei?

- En quin grau el nou servei millora la qualitat respecte el servei vigent?

- En quin grau el nou servei és accessible i fàcil d'utilitzar?

- Existeixen altres requeriments o recomanacions per part dels espectadors a tenir en compte a l'hora d'aplicar el servei?

Consentiment Informat

Avaluació subjectiva de serveis emergents d'accessibilitat: Audiodescripció en sistemes de TV digital

Jo, _____

• He llegit el full d'informació que se m'ha lliurat.
• He pogut fer preguntes sobre l'estudi.
• He rebut suficient informació sobre l'estudi.

• Comprenc que la meva participació és voluntària.
• Comprenc que em puc retirar de l'estudi:
1.Quan vulgui.
2.Sense haver de donar explicacions.

Presto lliurement la meva conformitat per participar en l'estudi.

Firma del participant

Data:

Instruccions dels escenaris

INSTRUCCIONS PELS PARTICIPANTS. ESCENARI 1

DESCRIPCIÓ

En aquest text es descriu en què consisteix la prova que realitzarà a continuació.

Li recordem que aquestes proves son per avaluar nous serveis que ofereixen continguts amb audiodescripció per a persones cegues. Les audiodescripcions (AD) descriuen tot el que passa en les escenes durant els moments en que no hi ha diàleg, de manera que les persones cegues es facin una idea dels continguts.

Al final de cada prova, haurà de completar un qüestionari referent a diferents aspectes del servei com, per exemple, la facilitat d'ús, el grau d'utilitat que creu que té, o la qualitat de les imatges i el so.

....

INSTRUCCIONS

En aquesta prova visualitzarà els continguts del serial "La Riera" de TV3. Aquest serial ofereix la possibilitat d'escoltar una versió amb audiodescripció per a cecs.

Visualitzarà els continguts de la sèrie a través del sistema de TDT (el que utilitza normalment a casa seva), i el nou servei que estem avaluant (streaming internet TV).

En cada cas, se li indicarà quin sistema està utilitzant. Visualitzarà la sèrie durant uns minuts per cada sistema.

Recordi que les preguntes del qüestionari faran referència a

- la qualitat de la imatge i l'àudio,

- a la facilitat a l'hora d'utilitzar el servei,

- a la utilitat del servei.

Per tant, és molt important que estigui atent a aquests aspectes durant la prova.

Tanmateix, una altra tasca serà la de sintonitzar l'emissió de la sèrie amb audiodescripció en el cas del nou servei. Per aquesta tasca, se t'informarà de com procedir en el seu moment.

Gràcies per la seva col·laboració.

INSTRUCCIONS PELS PARTICIPANTS. ESCENARI 2

DESCRIPCIÓ

En aquest text es descriu en què consisteix la prova que realitzarà a continuació.

Li recordem que aquestes proves son per avaluar nous serveis que ofereixen continguts amb audiodescripció per a persones cegues. Les audiodescripcions (AD) descriuen tot el que passa en les escenes durant els moments en que no hi ha diàleg, de manera que les persones cegues es facin una idea dels continguts.

Al final de cada prova, haurà de completar un qüestionari referent a diferents aspectes del servei com, per exemple, la facilitat d'ús, el grau d'utilitat que creu que té, o la qualitat de les imatges i el so.

....

INSTRUCCIONS

En aquesta prova visualitzarà els continguts del serial "La Riera" de TV3. Visualitzarà els continguts de la sèrie a través del sistema de TDT (el que utilitza normalment a casa seva).

En aquesta prova volem avaluar la possibilitat de que una persona escolti l'audiodescripció a través d'uns auriculars, possibilitant així que altres persones escoltin la versió estàndard al mateix temps (per exemple, una família on hi hagi un membre amb discapacitat visual).

La seva tasca serà sintonitzar i escoltar l'audiodescripció amb auriculars mentre visualitza la sèrie.

Les preguntes del qüestionari faran referència a

- la qualitat de l'àudio,

- a la facilitat a l'hora d'utilitzar el servei,

- a la utilitat del servei.

Per tant, és molt important que estigui atent a aquests aspectes durant la prova.

Gràcies per la seva col·laboració.

INSTRUCCIONS PELS PARTICIPANTS. ESCENARI 3.1.

DESCRIPCIÓ

En aquest text es descriu en què consisteix la prova que realitzarà a continuació.

Li recordem que aquestes proves son per avaluar nous serveis que ofereixen continguts amb audiodescripció per a persones cegues. Les audiodescripcions (AD) descriuen tot el que passa en les escenes durant els moments en que no hi ha diàleg, de manera que les persones cegues es facin una idea dels continguts.

Al final de cada prova, haurà de completar un qüestionari referent a diferents aspectes del servei com, per exemple, la facilitat d'ús, el grau d'utilitat que creu que té, o la qualitat de les imatges i el so.

....

INSTRUCCIONS

En aquesta prova es vol comprovar la viabilitat del servei que permet obtenir continguts amb audiodescripció sota demanda de l'espectador.

La seva tasca consistirà en accedir i visualitzar als continguts amb audiodescripció mitjançant l'aparell de televisió.

Els continguts que es seleccionen es visualitzen a la TV a la vegada que es van descarregant.

Les preguntes del qüestionari faran referència a

- la qualitat del vídeo i de l'àudio,

- a la facilitat a l'hora d'utilitzar el servei,

- a la utilitat del servei.

Per tant, és molt important que estigui atent a aquests aspectes durant la prova.

Gràcies per la seva col·laboració.

INSTRUCCIONS PELS PARTICIPANTS. ESCENARI 3.2.

DESCRIPCIÓ

En aquest text es descriu en què consisteix la prova que realitzarà a continuació.

Li recordem que aquestes proves son per avaluar nous serveis que ofereixen continguts amb audiodescripció per a persones cegues. Les audiodescripcions (AD) descriuen tot el que passa en les escenes durant els moments en que no hi ha diàleg, de manera que les persones cegues es facin una idea dels continguts.

Al final de cada prova, haurà de completar un qüestionari referent a diferents aspectes del servei com, per exemple, la facilitat d'ús, el grau d'utilitat que creu que té, o la qualitat de les imatges i el so.

....

INSTRUCCIONS

En aquesta prova es vol comprovar la viabilitat del servei que permet obtenir continguts amb audiodescripció sota demanda de l'espectador.

La seva tasca consistirà en accedir als continguts amb audiodescripció mitjançant un ordinador.

És necessari descarregar completament els fitxers amb els continguts per a poder-los visualitzar.

No obstant, en aquest cas no caldrà que avaluï la qualitat dels continguts.

Les preguntes del qüestionari faran referència a

- la facilitat a l'hora d'utilitzar el servei,

- la utilitat del servei.

Per tant, és molt important que estigui atent a aquests aspectes durant la prova.

Gràcies per la seva col·laboració.

INSTRUCCIONS PELS PARTICIPANTS. ESCENARI 3.3.

DESCRIPCIÓ

En aquest text es descriu en què consisteix la prova que realitzarà a continuació.

Li recordem que aquestes proves son per avaluar nous serveis que ofereixen continguts amb audiodescripció per a persones cegues. Les audiodescripcions (AD) descriuen tot el que passa en les escenes durant els moments en que no hi ha diàleg, de manera que les persones cegues es facin una idea dels continguts.

Al final de cada prova, haurà de completar un qüestionari referent a diferents aspectes del servei com, per exemple, la facilitat d'ús, el grau d'utilitat que creu que té, o la qualitat de les imatges i el so.

....

INSTRUCCIONS

En aquesta prova es vol comprovar la viabilitat del servei que permet obtenir continguts amb audiodescripció sota demanda de l'espectador.

En aquest cas, els continguts fan referència als arxius d'àudio que contenen les audiodescripcions (és a dir, podcast –només fitxers d'àudio).

La seva tasca consistirà en accedir als podcast amb audiodescripció mitjançant un ordinador. És necessari descarregar completament els fitxers per a poder-los escoltar. No obstant, en aquest cas no caldrà que avaluï la qualitat de l'àudio.

Les preguntes del qüestionari faran referència a

- la facilitat a l'hora d'utilitzar el servei,

- la utilitat del servei.

Per tant, és molt important que estigui atent a aquests aspectes durant la prova.

Gràcies per la seva col·laboració.

Photographs of the assembly

Photo 1: Making a meeting for scenario 2

Photo 2: Details of the application to select audio description

Photo 3: Details of the application for Scenario 3.1.
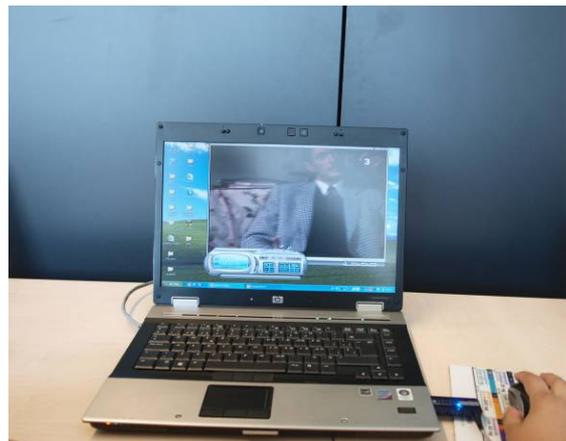
Photo 4: Scenario 3.2. Consumer PC

DTV4All Nutzertest: „Gebärdensprach-Dolmetscher auf hybriden Endgeräten"

Name Tester:
Name Dolmetscher:
Datum/Uhrzeit:

6. Fragen zu Mediennutzung:

6.1   Besitzen Sie einen Fernseher?
   ☐ Ja
   ☐ Nein

6.2   Wenn ja, wie oft nutzen Sie Ihren Fernseher?
   ☐ Nie
   ☐ Einmal im Monat
   ☐ Mehrmals in der Woche
   ☐ Einmal in der Woche
   ☐ Täglich

6.3   Nutzen Sie digitales Fernsehen? ***Sven: evt. erklären!***
   ☐ ja
   ☐ nein

6.4   Nutzen   Sie   eine   Set-Top-Box   oder   ein   Fernsehgerät   mit Internetanschluss?
   ☐ Ja
   ☐ Nein

6.5   Nutzen Sie einen Gebärdensprach-Dolmetscher im Fernsehen?
   ☐ Ja
   ☐ Nein

6.6   Bewerten Sie mit einer Schulnote 1-5 die Gestaltung des derzeitigen Gebärdensprach-Dolmetschers.
   ☐ 1

☐ 2
☐ 3
☐ 4
☐ 5

6.7 Haben Sie zu Hause einen Computer?

☐ Ja
☐ Nein

6.8 Wie oft nutzen Sie den Computer?

☐ Nie
☐ Einmal im Monat
☐ Mehrmals in der Woche
☐ Einmal in der Woche
☐ Täglich

6.9 Wofür nutzen Sie Ihren Computer?

☐ Internet
☐ Audio/Video
☐ Email
☐ Office/Schreiben
☐ Sonstiges: _____

6.10 Sehen Sie sich Videos im Internet mit   Gebärdensprach-Dolmetscher an?

☐ Ja
☐ Nein

6.11 Auf welchen Seiten?

_____

_____

_____

6.12 Sehen Sie sich Gebärdensprach-Videos im Internet an?

☐ Ja
☐ Nein

6.13 Auf welchen Seiten?

_____

_____

_____

7. Bewertung einer Test-Anwendung

<u>Anleitung für den Nutzer:</u>

Sie testen jetzt eine Anwendung. Diese Anwendung ist für so genannte hybride Fernsehgeräte gedacht. Das sind Fernseher, die auch einen Internetanschluss haben. Das ermöglicht, dass man eine Fernsehsendung mit eingebranntem Gebärdensprechervideo komplett über die Internet-Verbindung des Fernsehens erhält (on demand). (mit leichter Zeitverzögerung)

Wir stellen Ihnen dazu eine Aufgabe.

7.1 Benutzerfreundlichkeit und Barrierefreiheit der Anwendung

Aufgabe: Bitte schauen Sie sich jetzt die Tagesschau mit Gebärdensprach-Dolmetscher an.

	Ja	Mit Hilfe	Nein
2.1. Gebärdensprach-Dolmetscher aufrufen			

User Bewertung	Zu kurz	Richtig	Zu lang
Wie fanden Sie die Dauer der Texteinblendung zum starten des Gebärdensprach-Dolmetcher-Videos?			

User Bewertung	Unten	Oben
Wo wünschen Sie sich diese Texteinblendung?		

7.2 Allgemeines Feedback zur Gebärdendolmetscher Lösung

7.2.1 Was gefällt Ihnen an diesem Gebärdensprach-Dolmetscher-Dienst am besten?

_____

7.2.2 Was gefällt Ihnen an diesem Gebärdensprach-Dolmetscher-Dienst am wenigsten?

_____

8. Aufgaben zur Bewertung verschiedener Parameter

<u>Einleitung für den Nutzer:</u>

Sie haben eben eine Anwendung getestet, bei der eine Tagesschausendung inklusive Gebärdendolmetscher gänzlich über den Internet"kanal" on demand abgerufen wurde.

Die neuen TV Geräte mit Internetanschluss (hybrid) ermöglichen es aber auch, beispielsweise die Tagesschau ganz normal zu senden und dann auf dem Bildschirm mit dem getrennt über das Internet kommenden Gebärdensprachvideo zu vereinen. In diesem folgenden Test geht es um ihre Meinung: Wie sollte soetwas gestaltet werden?

8.1   Synchronität

<u>Erläuterung für Gebärdensprach-Dolmetscher:</u>

Die Internet-Übertragung des Gebärdensprachdolmetschervideos ist tendenziell Schwankungen unterworfen. Dementsprechend kann nicht von einer Synchronität beider Videoströme (Fernsehen und separates Gebärdensprachdolmetschervideo) ausgegangen werden. Wir möchten hier testen in welchem Ausmaß eine Verzögerung des Gebärdendolmetschervideos noch akzeptabel ist für die Tester. Ausgangspunkt ist hier die normale „Real-World" Verzögerung, die sich immer bei Übersetzungen ergibt.

Außerdem wollen wir noch zwei weitere Dinge testen:

Untertiteltests in DTV4All haben ergeben, dass die Tester es schätzen wenn die Untertitel gleichzeitig / simultan erscheinen oder sogar vorzeitig.

Entsprechend soll auch dies für die Gebärdensprecher getestet werden, So ergeben sich drei zu testende Szenarien:

> *1. Dolmetschervideo ist zum Hauptvideo verzögert.*

> *2. Dolmetschervideo gleichzeitig zum Hauptvideo.*

> *3. Dolmetschervideo ist dem Hauptvideo zeitlich voraus.*

8.1.1Dolmetschervideo ist zum Hauptvideo verzögert

8.1.2ACHTUNG: Stopp-Uhr!

<u>Erläuterung für Gebärdensprach-Dolmetscher:</u>

1. Szenario: Das Dolmetschervideo ist relativ zum Hauptvideo verzögert. Bis zu welchem Punkt ist die Verzögerung akzeptabel? Ausgangspunkt ist hier die normale „Real-World" Verzögerung, die sich

immer bei Übersetzungen ergibt. Beide Videos starten gleichzeitig. Eine wachsende Verzögerung wird simuliert, um die Akzeptanzschwelle herauszufinden.

Anleitung für den Nutzer:

Sie sehen einen Ausschnitt der Tagesschau mit Gebärdensprachdolmetscher. Bitte sagen Sie uns wenn Sie der Tagesschau nicht mehr folgen können.

User	Zeitpunkt
Zeitpunkt an dem der Nutzer nicht mehr folgen kann	

8.1.3 Dolmetschervideo gleichzeitig zum Hauptvideo

Erläuterung für Gebärdensprach-Dolmetscher:

Szenario 2: Das Dolmetschervideo ist simultan zum Hauptvideo, also „synchroner als in der „Real-World". Macht das aus Sicht der Nutzer Sinn?

Eine Simulierung der simultanen Situation wird dem Tester vorgespielt und von ihm beurteilt .

Anleitung für den Nutzer:

Es ist möglich, einen Gebärdensprachdolmetscher gleichzeitig zu senden. Sie sehen dazu jetzt einen anderen Ausschnitt der Tagesschau mit Gebärdensprachdolmetscher.

User Bewertung	gut	nicht gut
Wie finden Sie die gleichzeitige Übertragung?		

8.1.4 Dolmetschervideo ist dem Hauptvideo zeitlich voraus

8.1.5 ACHTUNG: Stopp-Uhr!

Erläuterung für Gebärdensprach-Dolmetscher:

Szenario 3: Das Dolmetschervideo ist dem Hauptvideo zeitlich voraus. Beide Videos starten gleichzeitig. Ein zunehmendes Vorauseilen wird simuliert, um die Akzeptanzschwelle herauszufinden.

Anleitung für den Nutzer:

Sie sehen jetzt noch einen weiteren Ausschnitt der Tagesschau mit Gebärdensprachdolmetscher. Bitte sagen Sie uns wenn Sie der Tagesschau nicht mehr folgen können.

User	Zeitpunkt
Zeitpunkt an dem der Nutzer nicht mehr folgen kann	

8.2 Bitrate (Qualität)

Erläuterung für Gebärdensprach-Dolmetscher:

Die Tester bekommen nacheinander das Gebärdenvideo in drei verschiedenen Bandbreiten (Qualitäten) präsentiert und geben ihr Urteil ob der Gebärdensprach-Dolmetscher in der jeweiligen Bitrate gut verständlich ist.

Anleitung für den Nutzer:

Sie sehen jetzt verschiedene Ausschnitte der Tagesschau mit Gebärdensprach-Dolmetscher. Bitte bewerten Sie nach jedem Ausschnitt die Qualität des Gebärdensprach-Dolmetsch-Videos.

Hinweis von Sven: Es geht hier nicht um den Inhalt, sondern nur um die Qualität des Bildes.

8.2.1 Bewertung der Bandbreite mit Bitrate 200 kbs >

User Bewertung	1 sehr gut	2 gut	3 mittelmäßig	4 schlecht	5 sehr schlecht
Bewerten Sie die Qualität des Gebärdensprach-Dolmetsch-Videos.					

8.2.2 Bewertung der Bandbreite mit der Bitrate 300 kbs >

User Bewertung	1 sehr gut	2 gut	3 mittelmäßig	4 schlecht	5 sehr schlecht
Bewerten Sie die Qualität des Gebärdensprach-Dolmetsch-Videos.					

8.2.3 Bewertung der Bandbreite mit der Bitrate 450 kbs

User Bewertung	1 sehr gut	2 gut	3 mittelmäßig	4 schlecht	5 sehr schlecht
Bewerten Sie die Qualität des Gebärdensprach-Dolmetsch-Videos.					

8.3 Bildgröße des Gebärdensprechers

Hinweis von Sven: Beachten Sie, dass die Größe des Fernsehbildes dabei auch beeinflusst wird.

Erläuterung für Gebärdensprach-Dolmetscher:

Die Tester bekommen nacheinander folgende Größen für das Gebärdensprachdolmetschervideo präsentiert und beurteilen für jede Größe mit Schulnoten ob die jeweilige Ansicht verständlich und angenehm ist.

        1. Grösse klein: 200x173

        2. Grösse (Phoenix) mittel: 300x260 (16:9)

        3. Grösse gross: 480x417

Anleitung für den Nutzer:

Sie sehen jetzt verschiedene Ausschnitte der Tagesschau mit Gebärdensprach-Dolmetscher. Bitte bewerten Sie nach jedem Ausschnitt die Größe des Gebärdensprach-Dolmetsch-Videos.

8.3.1 Bewertung der Bildgröße „klein" (200x173)

User Bewertung	1 sehr gut	2 gut	3 mittelmäßig	4 schlecht	5 sehr schlecht
Bewerten Sie die Größe des Gebärdensprach-Dolmetsch-Videos.					

8.3.2 Bewertung der Bildgröße (Phoenix) „mittel" (300x260 (16:9))

User Bewertung	1 sehr gut	2 gut	3 mittelmäßig	4 schlecht	5 sehr schlecht
Bewerten Sie die Größe des Gebärdensprach-Dolmetsch-Videos.					

8.3.3 Bewertung der Bildgröße „groß" (480x417)

User Bewertung	1 sehr gut	2 gut	3 mittelmäßig	4 schlecht	5 sehr schlecht
Bewerten Sie die Größe des Gebärdensprach-Dolmetsch-Videos.					

8.4 Position des Gebärdendolmetschers auf dem Bildschirm

<u>Erläuterung für Gebärdensprach-Dolmetscher:</u>

Die Tester bekommen nacheinander 12 Bilder (Screenshots) mit verschiedenen Positionen des Gebärdensprach-Dolmetscher-Videos gezeigt und beurteilen jede Variation mit Schulnoten.

<u>Anleitung für den Nutzer:</u>

Sie sehen jetzt 12 verschiedene Bilder der Tagesschau mit Gebärdensprach-Dolmetscher. Auf jedem Bild sehen Sie das Gebärdensprach-Dolmetsch-Video anders positioniert. Bitte bewerten Sie nach jedem Bild die Position des Gebärdensprach-Dolmetsch-Videos.

8.4.1Bewertung der Position des Gebärdensprechers

User Bewertung	1 sehr gut	2 gut	3 mittelmäßig	4 schlecht	5 sehr schlecht
Bewerten Sie die Position: Überlappend/Oben/Rechts					
Bewerten Sie die Position: Überlappend /Mitte/Rechts					
Bewerten Sie die Position: Überlappend /Unten/Rechts					
Bewerten Sie die Position: Überlappend /Oben/Links					
Bewerten Sie die Position: Überlappend /Mitte/Links					
Bewerten Sie die Position: Überlappend /Unten/Links					
Bewerten Sie die Position: Neben dem Hauptvideo/Oben/Rechts					
Bewerten Sie die Position: Neben dem Hauptvideo /Mitte/Rechts					

Bewerten Sie die Position: Neben dem Hauptvideo /Unten/Rechts					
Bewerten Sie die Position: Neben dem Hauptvideo /Oben/Links					
Bewerten Sie die Position: Neben dem Hauptvideo /Mitte/Links					
Bewerten Sie die Position: Neben dem Hauptvideo /Unten/Links					

9. Fragen zur Meinung der Nutzer:

9.1 Grundsatzfragen

9.1.1 Würden Sie generell solche neuartigen Gebärdensprach-Dolmetscher-Dienste am Fernsehgerät nutzen?

_____

Es gibt verschiedene Möglichkeiten, einen neuartigen Gebärdendolmetscher für digitale Fernsehgeräte mit Internetanschluss anzubieten. Was bevorzugen Sie persönlich?

    1. ☐ Gebärdensprecher, der mit der Sendung „on demand" komplett über den Internetkanal auf den Fernseher kommt (die erste Anwendung, die Sie getestet haben).

2. ☐ Ein zuschaltbares Gebärdendolmetschervideo: Hier wird das Gebärdendolmetschervideo über die Internetverbindung übertragen und wird live mit der laufenden Sendung kombiniert.

9.2 Wie wichtig ist es Ihnen, selber die Position und die Größe des Gebärdendolmetschers bestimmen zu können?

☐ sehr wichtig

☐ wichtig

☐ unwichtig

9.3 Was finden Sie am besten?

1. ☐ Dolmetschervideo ist zum Hauptvideo verzögert – so wie auch in der normalen Welt oder auch jetzt bei Phoenix

2. ☐ Dolmetschervideo läuft genau gleichzeitig zum Hauptvideo

3. ☐ Dolmetschervideo ist dem Hauptvideo zeitlich voraus

9.4 Haben Sie noch abschließende Bemerkungen zur barrierefreien Gestaltung eines neuartigen Gebärdendolmetscher-Dienstes für den Fernseher?

_____