COMPETITIVENESS AND INNOVATION FRAMEWORK PROGRAMME

ICT Policy Support Programme (ICT PSP)

ICT PSP call identifier:	**ICT PSP/2007/1**
ICT PSP Theme/ Objective identifier:	**2.1: Accessible digital Audiovisual (AV) system (Pilot B)**
Project acronym:	**DTV4All**
Project full title:	**Digital Television For All**
Grant Agreement no.:	**224994**
Contract duration:	**01.07.2008 – 31.03.2011**
Deliverable no.:	**D3.7 / D4.7**
Deliverable title:	**Joint Recommendations for Future Access Services, Devices and Platforms**
Nature of deliverable:	**Report**
Dissemination level	**Public**
Due date:	**28.02.2011**
Actual delivery date:	**06.05.2011**

Document title and history: **D3.7 – Recommendations for Future Access Services, Devices and Platforms**			
Version no:	*Date:*	*Changes*	*Contributor*
001	24-09-2010	First draft distributed for discussion	IRT
002	27-09-2010	Revisions to the first draft	Brunel
003	11-10-2010	Contribution by TVC	TVC
004	19-11-2010	Revisions to the first draft	IRT
005	09-12-2010	2nd revision	IRT
006	13-12-2010	Editorial revisions	Brunel
007	14-12-2010	Revisions to the second draft	IRT
008	20-12-2010	Final edits to section 2.4 (TVC)	TVC
009	21-12-2010	Revisions	rbb
010	07-01-2011	Editorial work	IRT
011	07-01-2011	Additions	rbb
012	21-01-2011	Editorial revisions	Brunel
013	27-01-2011	Additions	rbb
014	01-02-2011	Editorial revisions	Brunel
015	02-01-2011	Additions	IRT

Document authors / organization:	**W. Brückner**
Partners contributed:	**All**
Availability:	**Drafts**
Circulation:	**Partners**

Table of Contents

1. Executive Summary

This deliverable provides recommendations for future access services based on the results derived from the laboratory tests with pertinent user groups carried out in the DTV4All project. These laboratory tests are described in detail in the deliverables D3.2 – D3.6. The basis of these laboratory tests were so called demonstrators.

These demonstrators called "Emerging Access Service Demonstrators" encompass Video Signing, Clean Audio, Reduced Playback Speed, Enhanced Audio Description, Enhanced Text Services and Text-to-Speech Applications.

The current document elaborates the conclusions which can be drawn from experiences gained through laboratory testing of each individual demonstrator taking into account the respective test set-ups and methodological approaches.

This deliverable also deals with how the knowledge gained in this project could be applied in the real broadcast world.

Sections 2 analyses the results obtained from laboratory tests of the demonstrators. For each demonstrator individually tailored recommendations are given.

In Section 3, an attempt is made to compare the benefits of the different access services and to give generic recommendations for future application within the European Union with respect to implementation and dissemination issues.

Finally, Section 4 provides some web references where further information can be found in relation to these services.

2. Specific Recommendations

This section focuses on the test results gained from laboratory tests making use of the latest techniques for barrier-free Digital Television and the conclusions which can be drawn from these tests.

It can be stated that all the demonstrators more or less fulfilled the needs of different user groups suffering from different disabilities: hearing impairments, visual impairments, and cognitive impairments. Because of the lack of homogeneity of the user groups there is also a lack of homogeneity in the test results. This makes the test results hard to compare. Generally speaking the benefits of each emerging access service have to be carefully considered and compared with the (estimated) efforts required for the implementation of the service. In most cases effort equates to cost. To compare the effort versus the benefit is an adequate solution to the challenge of providing a transparent presentation of the results gained from the laboratory tests.

The efforts associated with providing a particular access service are sub-divided into the extra efforts it requires on the side of broadcasters and on the side of the manufacturers. It might be the case that there will be no or little extra efforts associated with providing a service on the broadcaster's side but relatively great effort on the receiver side and vice versa. As a rough guide, the costs/efforts are identified as being none, low, medium or high. This coarse scale reflects the test results and will be sufficient for the needs of most of the readers of this document. In conjunction with the efforts or cost for a dedicated emerging access service conclusions are provided concerning the benefits for the end-users which have been deduced from the field tests and laboratory tests carried out in the project. As stated above, these users are not a homogeneous group of people even if they suffer, broadly speaking, from the same kind of disability, e.g. reduced hearing acuity. It is a matter of fact that what one hearing impaired person rates of high value, another may rate low, because the two suffer from different types or degrees of impairment. However, the fact that emerging services can offer personalisation features opens up opportunities to address the issue of the heterogeneity of impairments. Concerning sight impaired users HbbTV-based text applications are a good example of this. As elaborated in D3.5, such new text applications allow the colours of the screen background and the text to be adjusted for better contrast. In DTV4All all possible variants were tested and seven variants were used after the test. These seven colour

variants cover the needs of a representative user group in terms of utmost contrast. Even a test person suffering from total colour blindness benefited from the customisable colour variants. He chose white text on a black background for utmost contrast which helped him to read better. Of course, there are limits: The fact that the portal for accessing HbbTV applications uses colour buttons for quick navigation was of no use to this particular test person because he simply cannot recognise colours. Here, further improvements would need to be done, i.e. remote control and application designs which work with shapes instead of colours. Another opportunity for addressing heterogeneity through personalisation is presented by sign translator videos delivered to so called hybrid devices with an Internet connection. rbb tests clearly showed that signing deaf persons have different needs as to the size and position of the signing video. Customization like the option to enlarge the signing video especially helped the group of Usher patients for whom deafness is accompanied by sight impairment. There were test persons who disagreed for various reasons with the majority of the test persons in a specific test. The reasons for such judgements can be found in particular test persons having specific disabilities for which a specific test environment did not meet the needs of the individual concerned. For instance, there exist several dialects of the sign language within a given country. A test for sign language in DTV may contain some gestures which a particular test person may simply not understand.

The recommendations given in the following sections should be regarded as a first approach to a rather complex set of problems. Moreover, the recommendations will give important indications as to how to improve existing access services or even help make the decision perhaps to introduce a completely new access service. DTV4All may give guidelines with respect to each access service considered but will not relieve implementers of access services of the need for further enquiry into particular services.

In the long access services will develop to address the plethora of individual needs. The results gained from the laboratory tests in DTV4All and the conclusions drawn from these tests show a clear and demand for personalised access services on part some of the test users and show that the emerging services which were tested are leading in the right direction and would only need minor modifications before being entered into regular operation. However, additional work is needed to identify further opportunities to address the heterogeneity of impairments and to make user groups aware of the possibilities for personalisation (customisable subtitles in terms of font and colour, customisable signing solutions for truly hybrid solutions etc).

2.1 Recommendations for Video Signing

2.1.1 Introduction

For many years broadcasters have provided sign language interpretation for a strictly limited number of programmes, in many European countries sign language interpretation of television programmes is not available. Where available the video image of a human sign language interpreter is usually superimposed on the broadcast video signal and transmitted with it. As the production costs of providing sign language interpretation are not insignificant and as the image of the video signer is displayed to all viewers, programming with sign language interpretation is typically limited to current affairs programmes aimed specifically at the signing deaf and a few news bulletins. To make sign language interpretation more widely available one of the challenges is to find solutions were a video signer can be displayed on demand, i.e. switched on and off with the main programme. In order to reduce the cost of providing a human sign language interpreter some research has strived for solutions where a computer generated character, an avatar, is animated to provide automatically sign language translations.

A first prototype for parallel delivery of a human sign language translator was done in the Framework Programme 5 project SAMBITS[1] using an additional video stream sent through the private section of an MPEG2-TS (TS: Transport Stream) that was superimposed over the main video at the Set-Top-Box. This was followed up in the Framework Programme 5 project SAVANT[2] which implemented parallel delivery of a sign language translator using an additional video stream sent over the Internet. This was superimposed over the main video received in a broadcast signal at the Set-Top-Box. As the delivery chain involved two different routes, a sophisticated synchronisation mechanism was developed and implemented both at transmission and at reception/presentation ends. This concept of synchronised parallel delivery was presented at IBC 2003 and 2004, IST Event 2004, at the 2004 IEEE International

[1] http://www.elec.qmul.ac.uk/mmv/sambits.html

[2] http://dea.brunel.ac.uk/project/Savant/

Conference on Multimedia and Expo, and under the aegis of the Framework Programme 6 Specific Support Action PARTAKE[3] in Beijing (China) in 2005.

Subsequently, in IBC 2009, BBC R&D (of UK) and NHK (of Japan) demonstrated a similar, but totally independent, synchronised parallel delivery system with a different middleware and a different buffering concept.

The system used for the above demonstrations between 2003 and 2005 was kept by IRT which was a participant in both SAVANT and PARTAKE. In 2009, IRT modified the 2005 version of the SAVANT system demonstrated in China, for the purposes of conceptual demonstration of emerging access services in Germany by rbb. See deliverables D3.3 and D3.5 for details of the synchronisation mechanism used, and deliverable D4.3 for details of the demonstration in IBC 2009..

For the use of RAI in Italian tests of emerging access services, a replica of the German 2009 version of the SAVANT system was configured by Brunel, the co-ordinator of both SAVANT and PARTAKE, and delivered to RAI for tests using new Italian language content made by RAI, see deliverable 3.6 for details of the results of these tests.

As the original software for the top-end (broadcaster side) synchronisation mechanism developed in the SAVANT Project was developed by another SAVANT partner, Brunel and IRT have the executable code and the right to use it for their own purposes under the consortium agreement of the SAVANT Project. However, Brunel and IRT do not have access to the source code, partly due to fact that the section in the partner that created the code was dissolved a few years after the end of the SAVANT Project. Furthermore, the receiver-end synchronisation mechanism is heavily dependent on middleware functions of IRT's edition of Multimedia Home Platform (MHP) which has not been used in commercial Set-Top-Boxes sold in Europe. These constraints severely restrict the modifications than can be made to the SAVANT system beyond those made for the purposes of DTV4All tests of emerging access services. In this context it is pointed out that the SAVANT head-end synchronisation mechanism could not be reverse engineered due to complications arising from revisions to the standards used by broadcasters such as those for metadata and Material Exchange Format (MXF) made since the SAVANT project ended. Furthermore, for the SAVANT receiver

[3] ftp://ftp.cordis.europa.eu/pub/ist/docs/ka4/au_fp6_partake_en.pdf

synchronisation mechanism synchronisation information is required including additional time stamps for presentation synchronisation. This overhead is not negligible.

However, methods are emerging on the market for the convergence of broadcast and Internet delivered video content. While these can enable the enhancement of broadcast programmes with content delivered over the Internet they also allow for simple switching between broadcast and Internet delivered television. Accordingly, DTV4All investigated the use of these existing methods for delivering sign language interpretation with a television programme by switching from the broadcast programme to an IP-based delivery of the programme with a sign language interpreter superimposed on the programme:

IRT and rbb developed an HbbTV-based application to this end. Two TV-signals were produced, one without the signer image and one with the signer video superimposed. While the first one is broadcast on air, the second one is delivered via the Internet. This approach alleviates the receiver from performing time synchronisation between the broadcast and the IP-signal as either one is selected by the hybrid set-top-box. Such hybrid boxes (e. g. HbbTV standard) have been available in the European market since 2010. A pertinent demonstration was given at IBC 2010, see deliverable D4.5. A HbbTV-based hybrid TV-based demonstrator was used for the expert user test of the emerging access service Video Signing, described in deliverable D3.6.

In Brunel, a switching solution was developed using a set of MHEG (Multimedia and Hypermedia Experts Group) functions and an on-demand IP stream server. In this implementation, the additional data rate added to that of the main broadcast content to enable a seamless switch to the same programme content with sign language interpretation is negligible. The additional delivery costs for the broadcaster are not high being limited to the switch function data, an on-demand IP stream server, and a broadband Internet connection. It is pointed out that the MHEG functions used to achieve the switching are available in the Internet enabled TVs on the European Market. Furthermore, functions comparable to the MHEG functions used are available in MHP.

RAI also tested the delivery of a signer to a small handheld mobile phone display, see deliverable D3.5, on a Personal Digital Assistant (PDA), and demonstrated the service at IBC 2010 in Amsterdam.

The test results from the Work Package 3 tests of emerging access services indicate that we may not have to synchronise the sign language interpreter video with the main video

as accurately as within one frame (1/24 or 1/25 sec) which is the accuracy required for lip synchronisation of dubbed sound. The users' demands are for synchronisation with accuracy of a few seconds either way. This is due to the nature of the sign language which has a different linguistic structure from natural spoken language. Under this condition, we do not need rigid synchronisation of the sign language interpreter video with the main video.

2.1.2 The effort or costs compared to the added value or benefit when using methods for switching between broadcast and broadband

The availability of sign language interpretation is an enormous help to its user community. The benefits of programming with sign language interpretation for the end-user are documented in earlier deliverables, see D3.6, page 13ff.

The majority of television viewers, however, do not understand sign language and could regard the presence of a sign language interpreter on the screen as annoying. For this reason, this service must become optional for users (a service that can be switched on and off). Hybrid set-top-boxes capable of decoding streamed media applications delivered through the Internet and displaying them on Digital Television screens are now on the market. The platforms available for enabling this include HbbTV and YouView. In principle broadcasters could make use of the Internet to deliver a programme with superimposed sign language interpretation on demand over the Internet. This solution was tested in DTV4All. For video distributions over the Internet the current compression standard is MPEG-4 which reduces the bandwidth required by a video of signer superimposed on a standard definition television frame to 1.75-2.25 Mb/s. Currently, this can be too high to use with HbbTV because the average household in Europe is still limited, especially in rural areas, to an average bitrate of approximately 2 Mbit/s. This is one of the reasons why German broadcasters currently restrict themselves to an average bitrate of approximately 1.5 Mbit/s for their video-on-demand content of standard definition[4]. However, in the medium term it is expected that the average bitrate available to the average household in Europe will increase significantly.

[4] Download of High Definition content is currently offered by the German public service broadcasters at data rates of 3,5 – 4 Mbit/s.

Concerning technical costs for the broadcaster such a solution is relatively cheap to provide. The server providing the service on demand can make use of the already established streaming infrastructure. Using the broadcasters existing web server means that extra costs of streaming a programme with sign language will be relatively low as long as the number of parallel video streams remains relatively low as should be the case for programmes with sign language interpretation. Production costs, however, will also involve resources for the studio recording of a human sign language interpreter which would increase the total cost for broadcasters from "low" to "high". The estimated extra costs at the receiver side are currently expected to be "low" because of the cumulative dispersion of hybrid set-top-boxes in the market. Furthermore, these costs are expected to go down in the near future, once hybrid broadcast broadband browsers are implemented in most or all set-top-boxes. This is however assuming that the user has an existing broadband connection faster than 2 Mbit/s. Summing up, the advantage for broadcasters and end users alike lies in the fact that services can be offered which may switched on or off on demand by the user(s) and thus do not irritate those who do not require access services and also in cost-efficient distribution.

2.1.3 Recommendations

Broadcasters offering sign language interpretation have to be aware that only a minority within the group of hearing impaired viewers are able to understand sign language. If they want to establish services offering sign language interpretation they could offer optional on demand services by providing sign language interpretation as an IP-based part of a hybrid TV service provided the members of the user group concerned have access to sufficiently fast Internet connections. Hybrid set-top-boxes and Integrated Hybrid digital television sets are available and can be made ready for use for this access service scenario. Independent of the technical means used for delivering the service, the tests involving deaf people did reveal a desire for personalisation with respect to the image size of the sign language interpreter, its colour design and position on the screen. This demand could be met by hybrid solutions. Generally, any solution has to allow proper recognition of all gestures and of the face expressions, and in certain countries the mouth and tongue movements of the person doing the signing. Furthermore, a

number of key recommendations for hybrid sign language interpretation services can be made:

1. A solution where a video on demand with a superimposed sign language interpreter is delivered via the IP-channel of a hybrid device has been shown to be feasible.

2. Access to such a solution must be easy. The pop-up window or icon announcing the availability of the on-demand service should remain on screen for several seconds. In future such a pop-up application could be replaced by a dedicated button on the remote control which automatically calls up any programme with sign language interpretation available on demand. Alternatively, on demand delivery of signed programmes could be entered into a user's personal profile on a hybrid TV system and would thus occur automatically each time it is available.

4. To allow for personalisation future solutions should enable a separate signing video to be delivered over the Internet that can be combined with the main broadcast programme on the screen. For this the hybrid device should be capable of decoding two different video streams in parallel. Concerning such a future "truly hybrid solution" DTV4All tests resulted in the following recommendations:

- The sign translation video should be customisable to be adjusted by the user in terms of size, position and design of the sign translation video.
- Deaf people welcome a congruent sign language translation. The matching can be improved through buffering a hybrid stream which overcomes the natural delay in real world sign language translation.

2.2 Recommendations for Clean Audio

2.2.1 Introduction

Clean Audio is a technique which enhances the dialogue of the main audio content of a broadcast. The level of the spoken words is enhanced with respect to any background sound present in the main audio sound track. If the Clean Audio cannot be created in the receiver, which is the case with products currently in the market, the Clean Audio has to delivered on an additional sound track. If original multi-track recordings including a dialogue only sound track are available or if dubbed imported material is to be broadcast, an alternative to broadcasting Clean Audio would be to broadcast a dialogue only track in its place. However, most hearing impaired users in the DTV4All tests at rbb considered this an inferior solution to Clean Audio that they found audible.

Clean Audio and dialogue only tracks are emerging access services that are expected to be a step towards improved intelligibility for large user groups of hearing impaired people despite the extreme heterogeneity of hearing impairments.

These access services must be supported by the broadcasters that have to provide an additional pair of sound tracks for stereo Clean Audio (or dialogue only) within the Digital Video Broadcasting (DVB) transport stream. This additional sound track is signalled in the standardized DVB Service Information (DVB-SI) in the same manner as Audio Description.

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

Successive laboratory tests were undertaken by rbb and UAB in late 2009 and spring 2010 with two groups of test persons. The test group at rbb comprised 18 test persons who were quite heterogeneous as to their degrees of hearing impairment and as to which assistive devices they used. The test group at UAB comprised 10 test persons 9 of whom had substantial hearing-loss and one had minor hearing-loss. It should be noted that Clean Audio cannot be used as a substitute for a hearing aid. The test persons were asked not to change hearing aids and head-sets but keep their “usual living room procedures” in order to obtain realistic results. Clean Audio is a service that could be helpful to many people with a slight or medium hearing impairment.

2.2.2 The effort or costs compared to the added value or benefit when Clean Audio is provided by the broadcasters

On the receiver side, almost all set-top-boxes and Digital Television receivers available in the market nowadays are able to switch, more or less comfortably, to alternative sound channels when the content is delivered via DVB-S, C or T (S; Satellite; C: Cable; T: Terrestrial). As a consequence of this, no extra cost will be imposed on the users. Currently this feature is used in providing audio description by some broadcasters (see Section 2.4).

The situation looks different at the broadcaster's side. To create a Clean Audio track from a main audio track, there is a need for highly sophisticated and costly equipment and skilled personnel able to use it. However, most new material comes with a separate dialogue only track. This offers the possibility of doing a separate mix to produce a clean audio track.

2.2.3 Recommendations

From the test results we learned that some hearing impaired persons did not benefit from Clean Audio provided during the DTV4All tests. Nevertheless, the test results confirm the need for Clean Audio services as a significant number of test persons would benefit from them. The comments of the UAB test group can be summarised as Clean Audio is a very good solution. Furthermore, a lot of broadcasting material is available with dialogue only tracks.

Further investigations would need to be done to optimise this kind of access service. On the other hand, the potential audience is high. It is estimated that 10 – 30 % of all viewers are hearing impaired to a certain degree. This proportion will increase in future because of the aging society. Consequently, a positive recommendation can be given for establishing such a service in the near future.

2.3 Recommendations for Reduced Playback Speed

2.3.1 Introduction

This access service allows the playback of a broadcast programme at reduced speed. Some PC-based software solutions for reduced playback speed were tested by IRT and applied to TVC footage. As a result of this process, a DVD was compiled which was used for laboratory tests at the facilities of UAB. The TV signal also contained subtitles. Eight native Spanish speakers participated in the experiment and watched while monitored with the help of an eye-tracker. Playback speed rates of 100%, 90%, 80% and 70% were tested.

2.3.2 The effort or costs compared to the added value or benefit when Reduced Playback Speed capability is provided by Consumer Electronics manufacturers

No cost will arise on the broadcaster side because no modification of the original content is needed. On the consumer side, the receiver must be equipped with a hard disk or similar facility to store the content. Digital television receivers or set-top-boxes with the capability of reducing the speed of broadcast content in real time are not yet on the market.

Further effort would have to be made by consumer electronics manufacturers to provide pitch correction for the playback at reduced speed. Nevertheless, in cases where the set-top-box is already equipped with a hard disk (as high-end consumer electronics products usually are) it should be relatively easy to implement the functionality for reduced playback speed for recorded items with the help of appropriate software.

2.3.3 Recommendations

The test results have clearly shown that reducing the play-out speed can have a positive effect on the comprehension of both the visual and the audio content.

Dyslexics and some people with cognitive impairments could benefit from receivers that allow the playback speed of the video/audio content of a program to be reduced. It was demonstrated in the project that the comprehension of highbrow content like scientific programmes etc. can be improved and conversations can become easier to understand. For hearing viewers proper pitch compensation needs to be applied for correct intelligibility and acceptance, for hearing impaired viewers slowing down play-out offers the possibility of providing more detailed sign language interpretation and subtitles with addition information. Slowing down play-out speed with suitable pitch compensation can currently be done but the computational requirements are high, hence it is difficult to implement on an ordinary set-top-box. It should also be noted that slowing the play-out down distorts timing markers in the broadcast so slowing live broadcasts requires some time tracking method such as the one available with TVAnytime.

For the above reasons it would be beneficial to some high-end consumer electronics manufacturers to equip future digital television receivers with such capabilities. However, more reception experiments should be carried out in order to study the possible differences (if any), in both visual perception and comprehension, between the different groups of viewers that could potentially benefit from slowed-down video: deaf and hearing impaired viewers, older people, cognitively impaired people and so on. Equally, it would be interesting to continue doing further research in order to establish better criteria for reducing playback-speed for accessibility, for example, determining the balance between the impact on programme viewing due to reduction of playback speed and the benefits of adding extra contents to slowed-down subtitles.

2.4 Recommendations for alternative ways of receiving Audio Description

2.4.1 Introduction

Audio Description (AD) is an additional sound track containing narration for blind and visually impaired people. When listening to a television program, people who are visually impaired or blind need audio description to help them fully appreciate the context of what they hear. Audio Description (AD) enhances accessibility of television programmes by allowing blind and visually impaired viewers to hear a description of what the characters on screen are doing; i.e. the action, body language, facial expressions, costume or scenery are described by means of narration.

In principle, AD is a mature service, already on offer in some European countries, via broadcaster mix or receiver mix. It is an emerging service when delivered to set-top-boxes via the Internet as the synchronization between the broadcast and the IP-channel is not implemented in hybrid set-top-boxes currently on the market.

AD can be delivered using different techniques: over Digital Video Broadcasting with the help of a second pair of audio streams within the digital multiplex (broadcast only solution) or with the help of a shared distribution, combining broadcast and broadband delivery (called a hybrid solution). Of course, the TV-signal, including AD, can be streamed over the web as WebTV.

Besides deploying a mature workflow for creating and publishing AD enriched content via Digital Terrestrial Television, TVC has also prototyped emerging ways of delivering these services (shown at IBC 2010), and several tests were undertaken by TVC and UAB in order to evaluate the benefits that end-users perceived from these emerging AD services. Mainly, these tests were focused on evaluating the usability, usefulness and quality of the AD services provided.

The delivery platform AD prototypes and the accompanying user tests have provided valuable information not only about the technical feasibility of new ways of receiving AD services, but also about the cost of implementing both mature and emerging AD services.

2.4.2 The effort or costs compared to the added value or benefit when providing AD

When providing broadcaster mix AD content for a programme, a new Audio Engineering Society (AES)/European Broadcasting Union (EBU) audio channel must be generated that merges the standard programme audio channel with additional auditory descriptions of the programme scenes. This additional audio channel must then be synchronised with the programme it describes by the broadcaster. In the case of receiver mix AD the audio delivered via the Internet contains only auditory descriptions of the programme scenes, the synchronisation and merging of this audio channel with the standard programme audio channel is done by the receiver.

The advantages of the receiver mix option are that less data needs to be transmitted and that the user could potentially mix the audio description with any one of the other audio channels accompanying the programme. For example, a French language film broadcast in Germany could potentially be broadcast with the original French language audio on one audio channel and German language dubbing of the film on another audio channel. German language audio description could then potentially be mixed with the French language audio channel, though in this case the actual matching of the audio description to the programme audio may not always be correct. A more significant use of the receiver mix capability would be to allow the audio description to be mixed with clean audio. However, despite the efforts of some broadcasters, mainly the BBC, and the collaboration of some set-top-box and integrated digital television manufacturers (e.g., Sony and Panasonic), only a few new receivers provide the AD receiver mix option. The technology is proven but manufacturers are not yet widely building it into the next generation of set-top-boxes and integrated digital televisions.

On the other hand, the broadcaster mix option is a cheaper and at the moment more universal[5] way to provide AD content. Digital Terrestrial Television (DTT) allows an audio channel to be used solely for audio description, as users can access this service by selecting the corresponding audio channel from the DTT receiver's audio menu. As a consequence, no costs will rise on the receiver side because most set top boxes provide

[5] There are some blind and partially sighted people who already own set-top-boxes or Integrated Digital Televisions that cannot provide AD through receiver mix but are able to receive broadcaster mix AD.

the capability to switch over to a second audio track. The drawback is that this solution implies less flexibility and requires greater bandwidth which means some additional costs for delivery.

In any case, independently of whether the broadcaster mix or the receiver mix option is used, the worst handicap that AD faces in spreading more widely is the high cost of providing it for the broadcaster. The pay rate for audio describers varies vastly country by country and depends on the personnel used, i.e. in house personnel or subcontractors. Audio description work is often subcontracted, although in a few cases some audio describers are volunteers who are not paid or just receive an honorarium. In any case, the costing of this activity is variable. It depends on the hourly rates of the subcontracted personnel, the time spent in preparation (normally, 2 or 3 previews are necessary), the cost of creating the script, the cost of describing, the cost of transporting the audio files to the broadcaster, and the cost of integrating the audio description inside the system. In the case of a film, it takes about 2 or 3 days to write the AD script, and one or two days to edit the audio.

For example, DTV4All partner Televisió de Catalunya (TVC) has developed and tested mature and emerging AD access services. The audio description is often achieved by setting up a parallel workflow[6] that supplies a new broadcaster mix audio channel, so that costs can be kept at minimum. The main features of this production process are that it is fully integrated in the existing workflow (AD content is remotely generated using software and the Internet and then automatically integrated), and that only two people are needed at the most. As no postproduction studios are needed, costs can be as low as €10 per minute of broadcasting. However, this is still far more expensive than the average cost of subtitling, which is 4€/minute.

---

[6] A typical workflow is as follows: Content (usually a film or a series episode) is transcoded to a low-quality video, with the content time code superimposed over the video, and stored in a system. The task of generating the AD is assigned to one of the external describers, who connects to a software tool, and can remotely download the afore-mentioned video because s/he has been previously authorized to do so. This professional previews the content several times, creates the AD script, and then generates audio files (one per segment of audio description) using the software tool. The same software tool sends the audio files back to the broadcaster, where they are kept in file servers. Here, a second person (this time, an employee of the Broadcast station) takes part in the process by means of controlling the quality of the broadcast AD, and by means of assuring that all audio descriptions are ready before they go on the air. When the content is being broadcast, the continuity matrix switcher sends play and stop commands (via a web service) to specific equipment that creates the broadcaster mix audio description channel. This equipment has as its input the ready to broadcast serial digital interface (SDI) signal and the audio files with the audio descriptions, and outputs the same SDI signal but his time with an additional EBU audio channel with the broadcaster mix AD.

Audio description can currently be done for live content. To big sport events, the broadcaster ORF regularly sends with its video team two types of reporter. One is the conventional TV commentator, the other reports more in the style of a radio commentator. The latter sound is used for live audio description of the sport event, e. g a football match. People with visual impairments switch their TV set to this "radio-type" sound channel.

If we want to evaluate the relationship between the effort and costs implied in providing an audio description service, and the benefits (and value) that this service brings, some considerations should be taken into account.

First of all, what is the percentage of population that would benefit from AD? The traditional paradigm considers that only certain segments of the population will be predisposed to consume audio description. Mainly this is visually impaired people, 1.77% of the population in Europe, 2.59% of the population of the world, according to World Health Organization[7], and some immigrants who want to learn the local language. It should be said that AD and subtitling are powerful tools when applied to language learning, and although the percentage of immigrants in Europe is in many countries higher than 10%, according to the United Nations report World Population Policies 2005[8], only a small percentage of immigrants use this service. However, as a communication[9] on e-accessibility in 2005 said, the spectrum of possible AD consumers is much wider:

> *"People with disabilities constitute about 15% of the European population and many of them encounter barriers when using ICT products and services. In certain cases, older people can be faced with similar problems. Accessible ICT products and services have now become a priority in Europe, due to the demographic shift: 18% of the European population was aged over 60 in 1990, while this is expected to rise to 30% by 2030".*

---

[7] Kuwait Medical Journal, March 2005: http://www.kma.org.kw/KMJ/Issues/mar2005/WHO-Facts%20Sheet-Mar-05.pdf. In the table on page 73: The European and world populations are respectively considered to be 877.9 million and 6,213.9 million; Europe is estimated to have 2.7 million blind people (36.9 million in the whole world), and 12.8 million people with low vision (globally, 124.3 million). As a consequence, the number of people in Europe with visual impairments was in 2005 of 15.5 million out of 877.9 million people (globally, it was 161.2 million out of 6,213.9 million people.

[8] http://en.wikipedia.org/wiki/Immigration_to_Europe. In 2005, some of the European countries that were above 9% of the population comprised of immigrants were Switzerland (25%), The Netherlands (20%), Latvia (19%), Estonia (15%), Austria (15%), Croatia (15%), Ukraine (14.7%), Cyprus (14.3%), Ireland (14%), Moldova (13%), Germany (12.3%), Sweden (12%), Belarus (12%), Spain (10.8%, 12.2% in 2009), France (10.2%), Italy (10% in 2010) and The United Kingdom (9%).

[9] [SEC(2005)1095]. http://eur-lex.europa.eu/LexUriServ/site/en/com/2005/com2005_0425en01.pdf

So, if we only consider the sight impaired and the blind, the target audience for audio description services could be something less than 5% of the total audience. However, if we consider more carefully, there are many more potential users of AD services. For example, older people are another segment *(note: this segment partly overlaps with those of immigrants and blind and visually impaired people*) of the population potentially interested in AD services because hearing acuity decreases sharply after the age of 65[10]. In fact, if we attend to the European guidelines, the social inclusion paradigm has changed. This new paradigm considers that 100% of the population should have the capability to access to AD services because AD services should not be measured by means of the law of supply and demand but in terms of social need the same way as parking lots for disabled people, or ramps in buildings, should be universally available regardless of what percentage of the population is disabled.

To have an approximate idea of the cost of universal provision of AD on Digital Television, let's take a look at TVC's TV channels. TVC broadcasts 7 TV channels. To calculate the costs of AD, we will not consider the satellite channel nor the high definition channel (TV3 HD) because there is no new content on these channels, we will not consider channel 3/24 because it is a 24-hour news channel and there is little that AD can do here, and we will also not consider TV3 Sports, because it is too new and is still in a testing stage. As a consequence, we will calculate the costs of audio describing the maximum possible amount of the content broadcast on 4 TV channels, 2 of which share the same frequency.

Channel	Live contents	1st time content	AD content	Subtitles
TV3	52%	50.2%	4%	47%
33D	18.7%	34.1%	1.6%	53.7%
CS3/3XL	0.9%	9.8%	4.9%	73.8%

Table 1 Percentages of live, new, and audiodescribed content on the 4 TVC channels considered during the first quarter of 2010

Needless to say, the channels do not have the same audience share[11]. Despite this, we will consider these channels as a whole, and assume that on average three quarters of the broadcast contents could be audio described, but that only 31.3% of the contents are

[10] http://en.wikipedia.org/wiki/Population_ageing. Asia and Europe are the two regions where a significant number of countries face severe population ageing in the near future. In these regions within twenty years many countries will face a situation where the largest population cohort will be those over 65 and the average age will be approaching 50.

[11] While the TV3 channel tops the audience lists in Catalonia with something less than 20% of the total television audience, other channels like C33, 3XL or 3/24 only have 1 to 3% share of the total audience.

broadcast for the first time the rest, are repeats, continuity elements, or advertisements. In other words, this means approximately 6,170 hours of new audio-describable content per year, for these three 24-hour TV channels. If we assume the aforementioned cost rate of 10€/minute, and that the fixed maintenance cost is negligible (about 1,000€/year per workstation of the AD system), the conclusion is that if the maximum amount of content that could be audio described was audio described (>6,000 hours per year), that would cost around 3.7 million euros per year. In other words, reaching an AD target population of between 2% and 35% of the 7.5 million people in Catalonia would cost in the worst case around 50 cents per citizen per year in Catalonia. This number is just an approximation but it gives us have an idea of the order of magnitude of the costs of AD. Half a euro per citizen per year might seem a low cost, but it is not. In fact, compared to the total operational costs, which are about 303M€ per year (including 128.6M€ of personnel costs), the annual cost of audio describing 6,170 hours of content would represent a considerable increase in operational costs and an important "portion of the cake".

To sum up, although AD is clearly a social need and its benefits for blind and sight impaired users are high, the estimated extra costs of providing it on the side of the broadcasters are judged to be rather high. This is partly because of the need for tools to establish an automatic workflow. However, most costs arise from staff costs due to an expensive price for each minute for producing AD content. The estimated costs on the receiver side are moderate to low, depending on whether the broadcaster mix option (no further costs) or the receiver mix option (slightly higher cost of the set-top-boxes is expected) is used. The benefit for blind and sight impaired end-users is high in both cases, as proven by the results of the user tests undertaken in the project, see deliverable D3.6. A compromise is therefore needed. Technology should evolve and become cheaper and help with this issue, but audio describing is a labour-intensive task and most of its costs come from personnel costs. As a consequence, while the percentages of audio described contents should definitively increase, they should not reach their ceiling (unlike subtitling), mainly because of their cost. The awareness and demand of the population for AD services will eventually establish the expected level of provision.

2.4.3 Recommendations

The switch-over to digital television across Europe by 2012 represents both a challenge and an opportunity for accessibility services. Internet TV is an additional way of receiving television programmes that not only allows selection of channels but also services on demand. Here we have to distinguish between IPTV and Internet-TV (also called "TV over the top" or WebTV). IPTV takes place in a managed network providing some guaranteed quality of service. Internet TV uses the world-wide web and has no guarantee of quality of service. For the moment, IPTV systems provide three types of services: live television, catch-up TV and Video on Demand. However, in the near future it will allow users to define what they watch on television, as well as when and how they do so. In the same way, users will have the chance to decide when and how to consume accessibility services. ADis one of the services that could be offered thanks to this technology to promote the inclusivity of access services.

One of the project partners defined and set up three different AD)scenarios to be tested in a user evaluation test. The first scenario “Live streaming Internet TV” emulates DTT broadcasting (the current system) with AD but by means of the Internet (using an IP channel). The second scenario “AD consumption: individually in a group situation” is a solution that has been developed for families or groups where one or more members want to consume AD, while the other members prefer services without AD. For this scenario AD is received through a laptop with DTT tuner, and the AD channel can be heard through headphones. The third scenario is “On demand AD” where users can download audio and video files in a portable format. These files contain the audio description audio channel and users will be able to reproduce these files in their portable devices.

In short, these prototypes showed that the emerging AD services that have been analysed in this trial are technically viable, but some aspects of the quality of the service must still 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 graphical user interface. The services’ usefulness also received user approval, 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 AD services have been given a high level of approval by their consumers, that the services are easily accessible, and that their quality is high enough so as to consider launching them.

2.5 Recommendations for Enhanced Text Services

2. . nt od ction

HbbTV services are based on the Consumer Electronics – HyperText Markup Language (CE-HTML) standard and are able to display mixed text, graphics and pictures, like the usual Internet websites. Sight impaired users can be provided with the means to adapt HbbTV services to their personal needs and preferences for better legibility and understanding.

In autumn 2009, the so-called HbbTV-based teletext service for the German nationwide ARD channel, Das Erste, was launched. In DTV4All, rbb and IRT adapted this service to include features for better accessibility especially for visually impaired users. In December 2009 an early version of the newly developed HbbTV service having barrier free features was tested by nine users with a range of visual impairments. The service was very well rated by the testers. For most, it was a real improvement compared to the classic teletext service.

Customisable settings for the font size and colour were found to be especially helpful. The majority of testers also liked the straightforward structure and navigation patterns familiar from the Personal Computer. The findings showed that for real accessibility it is essential to offer a range of personalised settings that can be adapted to suit the needs of a wide range of visual impairments. These should include font enlargement options and adjustable colour settings for a better contrast. In addition, it is helpful if the access to the settings is obvious and well placed. User feedback was fed into the redesigned current version of the ARD HbbTV teletext service.

2. .2 e e o t o costs com a ed to t e added a e o ene it en sin o en standa ds i e and id ecei e sets

HbbTV services now entering the market can be of some benefit for visually impaired people even if they have not been enriched with dedicated accessibility features: The user test of enhanced text services, (see deliverable D3.5), verified that even a HbbTV service designed for a mainstream audience improves readability and usability of the

service. The users straightaway praised the new design without having even started the accessibility options.

The user tests showed also that specific personalisation features for better accessibility which were designed in DTV4All really make a difference for visually impaired users. Such functionalities for changing font size and colour sets can be easily implemented by using different Cascading Style Sheets (CSS) as the core of HbbTV is HTML-based (HTML: HyperText Markup Language). Here a number of pre-defined CSS layouts can be provided by the broadcaster, each holding a dedicated font size and colour set, so that users can choose the one that is optimal for their sight abilities. Due to the fact that HbbTV applications are realised through the use of CSS layouts anyway, such an additional implementation and provision of a number of CSS can be made without significant extra efforts and costs. The precondition is the evaluation of colour sets and font sizes suited to the needs of visually impaired people. Essential recommendations for this were derived in the DTV4All user test (D3.5). The development or adaptation of HbbTV applications will then include user interface elements where users can select and access various styles and thus adapt the appearance of the application to their needs. Additionally, there is absolutely no need to replace the HbbTV enabled set-top-boxes already in the market, as decoding and rendering web pages accordingly to the provided scripts and styles is a fundamental HbbTV capability.

The estimated marginal extra cost for making HbbTV applications adaptable can be neglected compared to the total costs of providing HbbTV services. Actually, according to IFA 2010 announcements costs for receivers will even drop and it is expected that by the end of 2011 about 8 million HbbTV enabled set-top-boxes and receiver sets will have been sold in the German market.

2. .3 Recommendations

HbbTV is a standard for combined broadcast and broadband delivery to connected televisions and set-top-boxes. The adaption of the appearance of text, e.g. by changing the font size and the colour sets, is done in the underlying technologies like HTML and the Cascading Style Sheets and can be implemented easily and in a user-friendly way. In DTV4All a number of key recommendations for new text services can be defined:

1. A HbbTV-text service should provide help for sight impaired users by offering options for considerable enlargement of the fonts, graphics and menus and also by offering a number of different contrast and colour variants for the text (for details please refer to previous deliverables of WP3).
2. The access point to the barrier free settings must be obvious, well placed, of high contrast and large enough.
3. The barrier free personalisation page must be accessible in itself; the representation used and the language generally need to be clear and easy to understand.
4. At the personalisation page, every setting change must have a clear and logic feedback mechanism (for example by providing feedback by sound for the sight impaired).
5. A barrier-free configuration of the service should not disturb the navigation mechanisms.
6. If the concept of using teletext page numbers is kept the page number input windows should be enlarged.
7. The screen representations of the text service colour buttons (Red, Green, Yellow and Blue) must be as clear as possible. Button labels must have clear contrast to the background colour.
8. The colours red and green should not overlap generally.

In summary, and when considering all the recommendations which have been given in this section, there is a clear "go" for a suggestion towards a use of this technique for all three parties: the consumers, the broadcasters and the consumer electronics manufacturers.

2.6 Recommendations for Text-to-Speech Applications

2. . nt od ction

Text-to-Speech (TTS) applications belong to the group of Audio User Interfaces (AUI). It is a means by which the users interact with their digital receivers. The target group for this are sight impaired and blind users. DVB-SI data contain a lot of Electronic Programme Guide (EPG) information. This information covers the title, start and end time as well as abstracts of scheduled programmes. Such data form the input to a 'Talking EPG' and to 'AUIs' where text data is transformed into speech. Sight impaired and blind users can benefit from the voicing of not only EPG screen prompts but also of teletext news, channel selection options, or recording facilities. AUI's convert text to speech for the sight impaired with the help of "speech application programming interfaces" (SAPI) available for several platforms recently offering synthetic voices of a good quality. In the early days of "Text-to-Speech" (TTS) the voices sounded quite artificial and were hard to understand. Nowadays TTS quality has improved considerably and such applications are expected to be built into consumer equipment for Digital Television at an affordable price and with good sound quality. This will considerably assist navigation by hand-held remote control devices.

It is noted that:

- At the DTV4All workshop on Barrier-Free Digital Television Thursday 28 October, 2010, European Parliament, Frank Kamperman of Philips on behalf of Digital Europe demonstrated an Integrated Digital TV with Text-to-Speech Engine.
- At the International Telecommunications Union (ITU)-European Broadcasting Union (EBU) Joint Workshop on Accessibility to Broadcasting and IPTV ACCESS for ALL, Geneva, Switzerland, 23 – 24 November 2010 in cooperation with the EU project DTV4All, the Royal National Institute for the Blind (RNIB), United Kingdom, demonstrated Smart Talk, the world's first talking digital freeview set-top-box. Frank Kamperman gave a presentation on behalf of Digital Europe in which it was reported that Digital Europe were working on two specification documents: TTS functional specification which addresses the behaviour of Digital Television with TTS attached and a TTS interface specification which addressed

the connection of a TTS converter to a Digital Television. These documents were ready for submission to formal standardisation around end of 2010.

User tests in DTV4All were performed using the German speaking Ammec-device, a Set-Top-Box for DVB-S, T or C (S: Satellite; T: Terrestrial; C: Cable) which provides audio interfaces. The envisaged laboratory test focused not so much on validating the Ammec device as such. Instead, it focused on generally validating "Audio User Interfaces" (AUI) for TV sets / digital receivers provided with the help of a text-to-speech engine (TTS). The Ammec was used for the test because at the moment it is the only TTS-based set-top-box on the German market. From the concrete Ammec-related results, general conclusions concerning this topic were derived.

The tests focused on two aims:

General aim:

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

Specific aim:

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

(For details please see previous deliverables D3.2 p. 26 ff and D3.4 p. 18 ff and D3.5 p. 124 ff).

2.6.2 Table of effort/costs versus the added value/benefit

Modifications towards audio interfaces in Digital Television receivers do not necessarily need support by the broadcasters. It is obvious that this will mean higher cost for the consumer electronics manufacturers. On the other hand, there are more and more tools to be found in the open source community for free or for moderate Intellectual Property Rights costs.

Alternatively, Text-To-Speech applications (TTS) can be realized by the broadcasters by using middleware platforms such as Multimedia Home Platform (MHP) or HbbTV.

For instance, information delivered by text (subtitles, teletext content, or EPGs) could be offered in a audio format (MP3 or AAC) on the server farms of the broadcasters, although the usability of streamed audio in HbbTV is still unclear. In Section 2.5, it is assumed, that the estimated extra costs for HbbTV can be neglected compared to the total cost of the convergence process. This probably will be true for TTS as a part of the multimedia content for the access services. It can be assumed that TTS will percolate the technology of set-top-boxes and integrated digital televisions and that, in the long run, the extra cost of TTS will decrease steadily until it is integrated into the total cost of hybrid receivers.

2.6. eco endations

The DTV4All test results show very clearly that a TTS-based device (Set-Top-Box) providing a wide range of functionalities is very much desired by its target users. All tested users would use it, if affordable, and find it very important. All the tested functionalities were deemed “very important” or “important” by the testers. These include, apart from choosing TV-channels, the EPG. and radio-related functionalities as well as recording and cutting of TV and radio programmes, accessing read-out Teletext services and, a little bit less important, playing CDs and DVDs.

As a result, a strong recommendation for moving towards Digital Television receivers capable of offering speaking interfaces can be made.

A combination of both techniques, namely providing TTS engines in the Digital Television receivers and/or audio delivered by the broadcasters was not tested.

According to the DTV4All test results, the following recommendations can be made on TTS-based audio interfaces for digital TV:

1. A very clear and logically structured operational concept is required. This is especially important for blind users to improve orientation. The concept should be specifically tailored to the sensorial powers of blind users and their skills in virtual orientation and memorization. The menu structure needs to be consistent and uniform across all navigation levels with clear and simple conveying of features and functionalities.

2. The concept must be thoroughly communicated in parallel with the provision of the given TTS application through the use of context-sensitive help pages within the application and the provision of audio manuals.
3. There should be consistent and clear naming of all functions with no ambiguous terms being used.
4. The should be feedback mechanisms for any user action or input (e.g. through vibration, sounds, speech), i.e. Examples are:
    - sounds for success or even spoken announcements like "action accomplished"
    - page numbers are read out while they are being entered
    - A fault protection function like, for example, a "warning"-signal whenever further buttons are activated while the mute mode is activated.
        - "Do you really want to delete your timer-entry for programme xxxx, then please confirm with OK", etc….)
    - A recognisable sound symbolising winding the film back or forward.
    - A sound signal for low battery power.
5. High level responsiveness of the system is a requirement.
6. Intelligent and automatic restructuring of screen text content is required.
7. Beyond the "UK Digital TV Receiver Recommendations" by UK's Digital Television Group a TTS remote control must provide a wide emitting angle.
8. There need to be dedicated, clearly distinguishable, and sufficiently sensitive buttons on the remote control for the most important functionalities like:
    - A "Where am I?" button for TV channels as well as User Interface (UI) navigation
    - Separate volume controls for the audio interface and the TV,
    - Clearly distinguishable separate mute buttons for the audio interface and the TV
    - Direct page number input
    - A pause button for stopping and resuming the interface speech
    - Channel Up and Down buttons
    - A dedicated button for activating the audio description audio channel of a TV programme
    - Keypad buttons for any number inputs, supplemented by an automatic reading audio feedback

- A correction button for incorrect user inputs
- A dedicated EPG button
- A button for adjusting the speed of the AUI speech

9. Text or tables etc. should be voiced along single words or even letters
10. A personalisable TTS configuration (e.g. profiles like ‘beginner’ or ‘advanced user’).
11. There should be an audio function that announces the time.

3. Generic recommendations

3.1 Conclusions

The analogue switch-off provides two forms of opportunities to improve access to digital television for those with physical, mental or age-related impairments:

1. To extend the provision of existing mature access services to European countries that do not currently provide them.
2. To provide new kinds of access services known as emerging access services such as those described in this document.
3. To provide the capability to personalise access services to meet individual needs.

To ensure the challenge is addressed, DTV4All took action on identifying, assessing and promoting emerging access services.

3.2 Final remarks

The most valuable contribution DTV4All could make was to identify the enablers that will allow a core set of emerging access services to be offered in all European Union member countries now or in the near future.

These key emerging access services were identified and described in a series of deliverables. The devices and platforms needed to support them in terms of technological feasibility, their perceived value to their intended users, and their business model viability have been described and demonstrated in the project.

Recommendations have been provided to bodies representing stakeholders in the access service value chain on the basis of which these bodies can take appropriate action in relevant standardization bodies.

4. Bibliography

General:

http://www.psp-DTV4All.org

http://ec.europa.eu/information_society/activities/einclusion/index_en.htm

Video Signing:

http://www.signingsavvy.com/

http://aslbrowser.commtechlab.msu.edu/browser.htm

Clean Audio:

http://www.irt.de/index.php?id=29

Audio Description:

http://en.wikipedia.org/wiki/Audio_description

Enhanced Text Services:

http://www.hbbtv.org/

http://www.hbbtv-infos.de/

Text-to-Speech Applications:

http://www.microsoft.com/reader/de/downloads/tts.mspx

