D4.1

Version	1.50
Date	2014/12/18
Author	IRT
Dissemination status	PU
Document reference	D4.1

D4.1 – Pilot-B Progress Report

Grant Agreement nº:	621014
Project acronym:	HBB4ALL
Project title:	Hybrid Broadcast Broadband for All
Funding Scheme:	CIP-ICT-PSP
Annex I reference version:	2013/10/22
Project Duration:	2013/12/01 – 2016/11/01 (36 months)
Coordinator:	Universitat Autònoma de Barcelona (UAB)
Beneficiaries:	Rundfunk Berlin-Brandenburg (RBB) Institut fuer Rundfunktechnik GmbH (IRT) Rádio e Televisão de Portugal SA (RTP) Televisió de Catalunya SA (TVC) Schweizerische Teletext (TXT) Vsonix Gmbh (VSX) Fundación Centro de Tecnologías de Interacción Visual y Comunicaciones VICOMTECH (VIC) Screen Subtitling Systems Ltd (SCREEN) Holken Consultants & Partners (HC) People's Playground BV (PPG) Universidad Politécnica de Madrid (UPM)

This project is supported by funding from the 'Competitiveness and innovation framework programme 20017-2013' of the European Union.

Project no. 621014
HBB4ALL
Hybrid Broadcast Broadband for All

CIP- Pilot actions
Competitiveness and innovation framework programme 2007-2013

D4.1 – Pilot-B Progress Report

Due date of deliverable: 2014/11/30
Actual submission date: 2014/12/18

Start date of project: 2013/12/01
Duration: 36 months

Lead beneficiary for this deliverable: Institut fuer Rundfunktechnik GmbH (IRT)

Revision 1.50

Project funded by the European Commission within the Seventh Framework Programme		
Dissemination Level		
PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

History Chart

Issue	Date	Changed page(s)	Cause of change	Implemented by
0.10	2014/09/15	-	Document structure	IRT
0.20	2014/11/06	All	Version after first round of contributions	IRT, RBB, TVC, UAB
0.30	2014/11/21	All	Version after second round of contributions	IRT, RBB, UAB
0.40	2014/11/28	All	Integration of further partner contributions	IRT, RBB, TVC
0.50	2014/12/03	All	Additional details in partners contributions. Executive Summary added	IRT, RBB, TVC, UAB
0.60	2014/12/05	All	Pre-final revision / version for internal review	IRT, RBB, TVC, UAB
0.80	2014/12/07	All	Internal review	UAB
1.00	2014/12/15	All	Final Version	IRT
1.50	2014/12/18	All	Format	UAB

Validation

No.	Action	Beneficiary	Date
1	Prepared	IRT	2014/12/05
2	Approved	UAB	2014/12/15
3	Released	UAB	2014/12/18

Table of Contents

Figures

Tables

Executive Summary

The Hybrid Broadcast Broadband for All project (HBB4ALL) investigates accessibility services in the new hybrid broadcast-broadband TV (HbbTV) environment. HBB4ALL Pilot-B "Alternative audio production and distribution" addresses the use of dedicated audio-based access services to specifically support users who are hard-of-hearing (Clean Audio), have vision disabilities (Audio Description and Spoken Subtitles) but also the provision of additional audio channels via IP more generically to allow multi-language transmission (and potentially multiplatform support, e.g. on-demand online media libraries as well as HbbTV applications). This document presents the activities and intermediate results of Pilot-B within the HBB4ALL project for the first project period. It includes an overview of the goals, current status and progress of the pilot.

Service pilots are foreseen to match each of the Pilot-B objectives (Clean Audio, Audio Description, other languages and language learning). In each of the pilot regions (Germany – Berlin-Brandenburg, Spain – Catalonia) regional sub-pilots are planned. The sub-pilot in Germany will focus on a personalized Clean Audio service via IP for dedicated user groups. Most likely the sub-pilots in Spain will be carried out as a combined service pilot offering the functionality of all alternative audio services to the end user. During the operational phase later in the project, in each sub-pilot partners will carry out dedicated user tests for the specific target groups to validate the implementations and to obtain further feedback on the users' experience of the respective audio services.

As an overall technical aim, across the mentioned objectives, Pilot-B addresses the delivery of the additional audio tracks via IP to allow a personalised provision of such tracks and to free up datarate in the DVB broadcast channel. DVB has the technical requirements for playing different audio tracks synchronised with one broadcast video. The use of HbbTV opens up additional possibilities, using the broadband (IP) channel to address also smaller target user groups. As the various versions of the HbbTV specification (1.0, 1.5 and the upcoming version 2.0) have different features with respect to the delivery of content/streams via IP, during the first project period the partners have addressed the various technical options to allow an optimal choice for the delivery of all additional audio services to HbbTV devices.

Two main technologies defined in HbbTV 1.5 and 2.0 can potentially be used for the services in Pilot-B: 1) MPEG-DASH to enable adaptive live streaming over the Internet and 2) multi-stream synchronisation to enhance DVB broadcast by synchronisation of additional audio stream(s) delivered via Internet. Due to a delay in the publication of HbbTV 2.0 and more importantly (resulting from this delay) the foreseen unavailability of end user devices supporting HbbTV 2.0 in the project's lifetime, the pilots will be based on the installed base of HbbTV 1.0/1.1/1.5 devices. Services cannot yet make use of specific HbbTV 2.0 features. Specifically for Pilot-B this means, that the "multi-stream synchronisation" feature will be missing, and thus a scenario with a live TV program (via DVB) enhanced by synchronisation of additional audio stream(s) delivered via IP (broadband) cannot be realised. Therefore Pilot-B will focus on on-demand services. HBB4ALL partners are looking into the possibility of realising a showcase based on early HbbTV2.0 prototype receivers.

During the first project period a variety of preparatory activities, supporting the foreseen pilots, have been carried out. Apart from the organisational preparations of the respective sub-pilots, the partners carried out various preliminary user tests, on the one hand to enhance the implementation of the Clean Audio generator and tune it for use in the pilots, and on the other hand to study novel Audio Description aspects and application possibilities.

The realisation of a Clean Audio service is a novelty; no actual service is on air at this moment. A lot of know-how is available amongst the project partners, as to how a Clean Audio signal can be generated and an implementation, tuned to the project's requirements, has been started. The technical approach for the generation of a Clean Audio signal has been adapted regarding the workflow requirements of the broadcasters and also based on the outcome of the first test with the Focus Group and subsequent lab tests in July 2014. The focus is now on using pre-mixed audio files as input signals, like 5.1 and stereo tracks, as these are readily available in existing audio production workflows. Preliminary user tests show that 5.1 content is the preferred input format. Also, it may be beneficial to distinguish between people wearing a hearing aid and people who are not and to adapt Clean Audio signals for either user group.

The Clean Audio generator will be implemented as a server-side solution. Thus, it is foreseen to provide several pre-produced Clean Audio versions with different mixing levels to the end-user, e.g. by allowing him to choose from different CA versions which are made accessible to him. This is a compromise between full personalization and technical feasibility and still supports the project goals.

Approaches have been identified on how the Clean Audio generator can be optimally integrated in the broadcasters' (RBB, TVC) production workflow and how the features may be offered to the end users (e.g. integrated in the existing players/applications for access services such as subtitles). Also the communication with the user associations was established as a basis for the further collaboration during the project.

Audio Description is being produced and transmitted as part of a DVB broadcast service by several broadcasters. TVC already produces Audio Descriptions using either spoken audio or in some cases Text-to-Speech. Similarly, additional audio tracks are being distributed via DVB at the moment. For HBB4ALL the idea is to enable this technology in the online space, specifically in the case of HbbTV by leveraging MPEG-DASH. To be able to make use of MPEG-DASH the distribution platform of service providers will need to be adapted, e.g. the generation of various audio components in the correct format, adaptation of the CDN via which content is delivered over the Internet. The requirements' work to this respect has been completed; currently the required content workflow changes are under analysis. Multiple audio support on the basis of MPEG-DASH is under development.

Preliminary user tests on the use of Audio Description as a didactic tool confirmed that such a service can enhance learning in general, and enhance learning foreign languages in particular. Another experiment, where Audio Description was offered on a secondary screen as part of the cinema experience for the blind and partially sighted, proved a high degree of user acceptance for this type of application.

The partners made good overall progress with respect to all pilot preparations (technically as well as organisationally). For the Clean Audio sub-pilots the details of the timing depend on the further development of the Clean Audio generator during the next months and especially on the outcome of additional tests early 2015. RBB/IRT foresee to start the sub-pilot phase in Germany/Berlin-Brandenburg in autumn 2015. Whether the sub-pilot will be an open pilot for all RBB viewers or a hidden test for a dedicated user group will be decided after the lab test as well. Based on the outcome of the test the decision will be made jointly with the production and programming department at RBB and the target group will be defined more accurately with support of specialists. Ideally, the content will be made available for several months in the RBB online media library.

TVC and UAB plan to implement tests with users in lab conditions of Clean Audio content in Spain. Depending on the lab tests results and quality of the enhanced audio it may be scaled up for a pilot. In case the open pilot is deployed, content will be published using TVC's existing HbbTV on-demand video application, *TV3A la carta*. At a later stage, content can also be reused for other platforms if feasible.

TVC plans to ideally offer the functionality of all alternative audio services combined in a single service pilot. To support this, internal prototyping of HbbTV VoD as well as tests of MPEG-DASH distribution of multiple audio tracks will be carried out during the next months in preparation of the operational phase. In parallel, work on end user requirements, also with respect to the automatic generation of Audio Description will be continued.

After the further preparatory work and planning of the operational phase (until mid-2015), the implementation of the services to be used in the sub-pilots will be defined (e.g. target platform(s), playout modality etc.), user groups will be fixed, timelines will be finalised. In autumn 2015 the start of the operational phase of Pilot-B is then foreseen; the exact timeframe for running the sub-pilots is currently under discussion.

1. Introduction

The Hybrid Broadcast Broadband for All project (HBB4ALL) investigates accessibility services in the new hybrid broadcast-broadband TV (HbbTV) environment. All media service providers, specifically broadcasters – both public and private – are required to increase the amount of programs supported by access services, and ensure the quality of the service. One of the most prominent challenges faced by all providers is to do this in a cost-efficient manner while at the same time offer the target users the opportunity to customise the access services they are using to best meet their personal preferences or needs. Also it is crucial that new services remain consistent with access services that are already available.

The use of Internet delivered audio-visual content and the new HbbTV application environment on connected TVs is seen as an opportunity to efficiently address these issues. Hybrid delivery platforms such as connected TVs and two-screen solutions enable a cost-efficient and convenient delivery of access services and allows specifically targeting also smaller user groups by providing such services, or parts of them, via Internet as an add-on to broadcast.

The HBB4ALL project will test access services in four interlinked pilots; Pilot-A: Multi-platform subtitle workflow chain; Pilot-B: Alternative audio production and distribution; Pilot-C: Automatic User Interface adaptation – accessible Smart TV applications; Pilot-D: Sign-language translation service. All pilots have started with a definition and preparation phase in December 2013.

HBB4ALL Pilot-B addresses the use of dedicated audio-based access services to specifically support users who are hard-of-hearing (Clean Audio), have vision disabilities (Audio Description and Spoken Subtitles) but also the use of additional audio channels via IP more generically to allow multi-language (and potentially multiplatform) transmission. The objectives of Pilot-B are addressed in more detail in section 1.2.

1.1. Purpose of the document

This document presents the activities and intermediate results of Pilot-B "Alternative audio production and distribution" for the first project period. It includes an overview of the goals, current status and progress of the pilot. It contains a summary of relevant information generated, collected and assessed within the time period covered by this report. Specifically, it presents intermediate results of many preliminary user tests, and the current plans for the operational phase of the service sub-pilots.

1.2. Summary of Pilot-B objectives[1]

As the name of Pilot-B already implies, there are several audio services (and hence target user groups) in our focus.

Especially for hearing-impaired people the dialog intelligibility of TV audio signals is a key criterion. Due to various reasons, the intelligibility of current TV audio mixes is often assessed as insufficient by many people, including elderly, non-native speakers as well as hearing-impaired people. Pilot-B therefore aims at enhancing dialogue intelligibility, as this would be beneficial for all these target groups. The HBB4ALL partners address the automatic generation and production of a respective Clean Audio signal, taking into account current audio production workflows, as well as the provision of a Clean Audio service to the end user by exploiting features of the HbbTV specifications. One of the service features will be to offer users the

[1] The original objectives were specified in the HBB4ALL Grant agreement for: CIP-Pilot actions Annex I -"Description of Work" (DoW).

possibility to adjust the dialogue intelligibility to their personal preferences (to a certain extent). With respect to an efficient delivery of the Clean Audio service to the end user it is foreseen to use the broadband (IP) channel.

Pilot-B also addresses people with vision disabilities, for whom currently already extra audio channels are being provided with the broadcast programs, containing a description of the action mixed with the dialogue (Audio Description (AD)). This technique allows vision-impaired users to follow what is going on in a far more effective way than by hearing the dialogue alone. To allow a cost-efficient production and delivery of such audio channels, first of all the automatic production (based on Text-to-Speech (TTS), using subtitles and pre-processed scripts as inputs) and potential improvements to this way of working are being addressed. Complementary to AD, also Spoken Subtitles will be considered. This service renders written subtitles in Spoken Subtitles, and they are a basic access service in "subtitling countries" where foreign production is broadcast in original language plus subtitles. With respect to the delivery of these signals, Pilot-B aims at using the broadband (IP) channel.

Last but not least, Pilot-B supports having different languages for a single programme. Given EU citizen mobility, TV content is not only seen by nationals, but also by large communities living away from home. There is a need to broadcast the same content in different languages synchronically. Also, such additional audio tracks may be used for language learning.

As an overall technical aim, across the mentioned objectives, Pilot-B addresses the delivery of the additional audio tracks via IP to allow a personalised provision of such tracks and to free up datarate in the DVB broadcast channel. DVB has the technical requirements for playing different audio tracks synchronised with one broadcast video. The use of HbbTV opens up additional possibilities, using the broadband (IP) channel to address also smaller target user groups. As the various versions of the HbbTV specification (1.0, 1.5 and the upcoming version 2.0) have different features with respect to the delivery of content / streams via IP, the partners will address the various technical options and select an optimal choice for the delivery of all additional audio services to HbbTV devices.

Table 1 summarises the objectives, indicates which regional sub-pilots are planned and which project partners are involved respectively.

Objectives		Partners by sub-pilot region	
		Germany	Spain
B-1	Personalized Clean Audio service via IP for certain groups (hard of hearing)	RBB IRT	TVC UAB
B-2	Automatically generated speech synthesis of Audio Description and Spoken Subtitles.		TVC UAB
B-3	Additional audio channels via IP as resources for multi-language transmission as well as language learning and acquisition		TVC UAB

Table 1. Overview of Pilot-B objectives and related sub-pilots.

1.3. Summary of Task 4.1

Task 4.1 "Pilot Definition and Preparation of Operational Phase" runs from M1 to M20.

In this task, the partners will analyse the State of the Art of relevant developments/technologies and the status quo of the audio services in broadcast. The requirements from end users' perspective will be evaluated, specifically by means of preliminary user tests (on specific features of the envisioned services) and surveys in cooperation with Focus Groups and associations representing the target users. The results will be taken into account during the further preparations of the service sub-pilots.

Also from broadcasters' perspective the requirements will be addressed, most importantly based on an analysis of the (audio) production and distribution workflows that are currently in place. The impact of the various distribution aspects for additional audio services (e.g. asset availability, publishing, metadata) will be considered when choosing implementations for the realisation of the audio services.

Various options for the generation and distribution/playout of the additional audio tracks will be considered. With respect to the multi-audio asset production, also the required metadata modifications will be accounted for to support handling of multiple audio assets on top of a video asset. To support a cost-efficient production, HBB4ALL aims at an automatic generation of CA and AD content and at using audio signals that are readily available in the audio production workflow.

Based on the information and evaluations, in task 4.1 will be decided in what form the alternative audio services will be implemented for the operational phase (task 4.3, starting in M20). Depending on the user requirements, targeted end user devices etc. requirements and features of the targeted services will be defined and the technical realisation of the services will be chosen. The implementation will depend on various criteria, e.g. which target platform (HbbTV version) must/can be supported, modality of the service offer (live, on demand, …) etc.; however, it will be chosen such that delivery to end-users and testers of the intended features is done without compromising overall results. Also, success criteria for user validation will be defined.

1.4. Summary of Task 4.2

Task 4.2 "Solution Integration and Trials" runs from M3 to M20.

This task will implement and integrate the technology and infrastructure needed for carrying out the operational phase (task T4.3). The (intermediate) results of task 4.1 will be used to make the appropriate selection of and adaptations to the components and to carry out the implementation of the Pilot-B services according to the required features and technological choices. The aimed audio services will be provided either on hybrid TV platforms, as additional DVB audio stream or as IP only service. Task 4.2 also will cover integration work to publish audio content, including publishing audio-visual assets and metadata in the appropriate formats, on the right platform etc.

In additional to the T4.1 preliminary user tests and surveys, expert or friendly user viewings e.g. with audio engineers or members of associations and federations of the hearing impaired and / or visually disabled will be conducted in order to gain feedback regarding technical feasibility, perceived user value of the audio services.

1.5. Acronyms and abbreviations

In this document, when necessary, identified partners within the project are referred to using the abbreviated names initially defined within the Consortium Agreement for HBB4ALL and reproduced on the cover sheet of this document. Abbreviations and acronyms are introduced in brackets in the text after the corresponding full text version.

1.6. Definitions and glossary

Access Service [UK] = **Accessibility service** [US] – The provision of additional services or enhancements that improve the accessibility of TV services for viewers with disabilities or special needs.

Accessibility – The degree to which a product, device, service, or environment is available to as many people as possible. Accessibility can be viewed as the "ability to access" and possible benefit of some system or entity. Accessibility is often used to focus on persons with disabilities or special needs and their right of access to entities, often through use of Assistive technology or Access Services.

Audio Description (AD) – Refers to an additional narration track intended primarily for blind and visually impaired consumers of visual media (including television and film, dance, opera, and visual art). It consists of a narrator talking through the presentation, describing what is happening on the screen or stage during the natural pauses in the audio, and sometimes during dialogue if deemed necessary.

Audio Introduction (AI) - Also known as introductory notes, show notes or programme notes – have been used in opera and theatre since the early days of AD. They are pieces of continuous prose, spoken by a single voice or a combination of voices lasting between 5 and 15 minutes. AIs aim to create a framework by which to understand the action; they have an information function providing relevant details such as running time, cast and production credits, as well as detailed descriptions of the locations, costumes and characters, and can convey a sense of visual style including camerawork and editing.

Audio Subtitling (AS) – Voicing subtitles. See Spoken Subtitles.

Audio-visual Content – All kinds of time-based content consisting of images and sounds.

Blind and Visually Impaired Patrons (B/VIP) – Total blindness is the inability to tell light from dark, or the total inability to see. Visual impairment or low vision is a severe reduction in vision that cannot be corrected with standard glasses or contact lenses and reduces a person's ability to function at certain or all tasks.

Clean audio (CA) – Enhanced audio signal by means of signal processing, with improved intelligibility of the dialogue with respect to ambient noise, “atmo”, music etc.

Digital Terrestrial Television (DTT) – Broadcast of land-based (terrestrial) signals. A terrestrial implementation of digital television technology uses an aerial to broadcast to a conventional television antenna (or aerial) instead of a satellite dish or cable television connection.

DVB – Digital Video Broadcasting a set of technical guidelines, standards and specifications to benefit and advance digital media markets world-wide. It was originally European in origin but today is an alliance of 250-300 companies.

HbbTV – Hybrid Broadcast Broadband TV is a major pan-European initiative building on work in the Open IPTV Forum aimed at harmonizing the broadcast and broadband delivery of entertainment to the end consumer through connected TVs and set-top boxes.

Impairment, age-related – A collection of sensory and cognitive impairments. In the general sense, it covers matters such as the deterioration of sight and hearing, memory impairment or memory loss. In the report, we look not only at persons who are elderly but also at the challenges facing children whose intellectual maturity has an impact on their ability to read subtitles. In principle, there can be other impairments that are related to stages in the person‘s life.

Impairment, hearing – A generic term including both deaf and hard of hearing which refers to persons with any type or degree of hearing loss that causes difficulty working in a traditional way. It can affect the whole range or only part of the auditory spectrum which, for speech perception, the important region is between 250 and 4,000 Hz. The term deaf is used to describe people with profound hearing loss such that they cannot benefit from amplification, while hard of hearing is used for those with mild to severe hearing loss but who can benefit from amplification.

Impairment, visual – Visual impairment (or vision impairment) is vision loss (of a person) to such a degree as to qualify as an additional support need through a significant limitation of visual capability resulting from either disease, trauma, or congenital or degenerative conditions that cannot be corrected by conventional means, such as refractive correction, medication, or surgery. The loss may cover visual acuity, significant central or peripheral field defects or reduced contrast sensitivity.

Metadata – Data about data, in our case information about television programs. This can be in the form of program listings or guides, or technical data delivered with the program to accomplish an access service.

MPEG-DASH or **DASH** – Motion Picture Expert Group – Dynamic Adaptive Streaming over HTTP. This technology is supported by HbbTV from version 1.5 onwards. It allows an adaptive streaming depending on e.g. the network capabilities.

Spoken Subtitles (SS) – The spoken rendering of the written (projected) subtitles or surtitles with a filmed or live performance. The subtitles can be read by a computerized voice (Text to Speech) or by a ‘voice talent’ or ‘voice actor’. This technique is mostly used in subtitling countries when broadcasting foreign production. Spoken Subtitles should not be confused with AD. They are complementary.

Speech Technology (ST) – Relates to the technologies designed to duplicate and respond to the human voice. They have many uses. These include aid to the voice-disabled, the hearing-disabled, and the blind, along with communication with computers without a keyboard. They enhance game software and aid in marketing goods or services by telephone. The subject includes several subfields: Speech synthesis, Speech recognition, Speaker recognition, Speaker verification, Speech encoding, Multimodal interaction.

Text to speech (TTS) – A type of speech synthesis application that is used to create a spoken sound version of the text in a computer document, such as a help file or a Web page. TTS can enable the reading of computer display information for the visually impaired person, or may simply be used to augment the reading of a text message. Current TTS applications include voice-enabled e-mail and spoken prompts in voice response systems. TTS is often used with voice recognition programs.

Video on Demand (VoD) – A system that allows users to select and watch video content of their choice on their TVs or computers. Video on demand is one of the dynamic features offered by Internet Protocol TV. VoD provides users with a menu of available videos from which to choose.

Voice-over (VO) – Also known as off-camera or off-stage commentary, is a production technique where a voice that is not part of the narrative is used in a radio, television production, filmmaking, theatre, or other presentations. It is placed over the top of a film or video and commonly used in documentaries or news reports to give explanations.

Voice-over-voice (VoV) – Typically used for non-native language content where a translation is mixed to the original signal. In contrast to "dubbing", where the original (non-native) voice is completely replaced by another (native) one, with VoV both voices are audible. If the relation (in dB) of the translation track and the original signal is too small, the intelligibility of the native language translation is considered as insufficient.

2. State of the Art and results from past projects

This chapter summarises relevant activities and results from work in the field of audio processing and speech intelligibility as well as related research on connected TV. Wherever possible, Pilot-B will build upon the respective results and recommendations to develop services and service components for the pilot.

Two projects at European level focused on relevant topics with respect to Pilot-B: DTV4All (section 2.1, covering Clean Audio, Reduced Playback Speed and Audio Description) and ADLAB (section 2.2). Audio Description (automatically generated) also was the focus of research in a Spanish Ministry research project FFI2012-31024 lead by UAB (section 2.3). Results of partners' experiences with Clean Audio and speech intelligibility (section 2.4) are summarised. The automatic generation of Audio Description is addressed in section 2.5. HbbTV-specific issues, related to delivery of multiple audio tracks with a TV program, are listed in section 2.6.

2.1.DTV4All

2.1.1. Clean Audio

The DTV4All[2] project implemented a CA service as a separate audio stream solution within the DVB multiplex. This CA signal was either a special mix made by an audio engineer or a result of several combined tools that were utilized to improve the stereo or 5.1 signal in terms of CA. The goal for both cases was to reduce the level of noise, music, atmosphere or any other signal which is not speech to a minimum. The results of the test group and the feedback of a wider audience were quite positive. However, both approaches cannot create CA automatically and the additional production costs avoid the integration of this service in the production chain of broadcasters.

HBB4ALL extends this approach by offering not just one CA stream via DVB but several, different CA streams via the broadband channel. More importantly, the HBB4ALL approach to generate a CA signal will be ready for automation, which increases the probability that broadcasters are willing to integrate such a service in their workflow.

[2] www.psp-dtv4all.org

2.1.2. Reduced Playback Speed

DTV4All also tested an approach to enhance the intelligibility of TV audio signals by reducing the playback speed of the audio/video content.

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. Playback speed rates of 100%, 90%, 80% and 70% were tested.

The test results clearly showed that reducing the playback 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 playback offers the possibility of providing more detailed sign language interpretation and subtitles with additional information.

However, as this technology has to be implemented in the TV or set-top box devices, this approach will not be followed in HBB4ALL.

2.1.3. Audio Description

Two different tests were taken on board regarding AD in DTV4All. The first tests were related to broadcast scenarios and were performed by TV3 and UAB. The second tests took care of how to produce AD content across Europe taking into consideration the different languages, language status, and language situation (monolingual, bilingual, etc.) and finally the translation modalities across Europe since they have direct effect on the AD (subtitle, dubbing, and voice-over).

Broadcast scenarios

Five emerging services for AD were proposed and evaluated. Users evaluated the usefulness, quality and usability of each service through questionnaires. Ratings and comments about the services were reported and analyzed. Figure 1 summarizes the results achieved for all the scenarios.

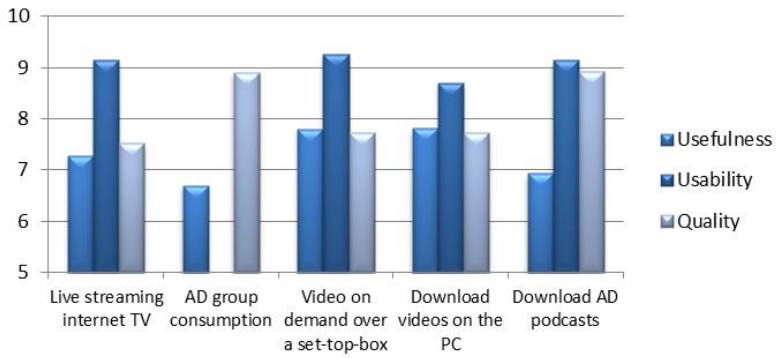

Figure 1. Summary of results for all the evaluated scenarios.

Although each AD scenario focuses on a specific use case, two general conclusions are extracted from the study: (1) users especially appreciate the fast access to video contents, and (2) the downloading of content for offline consumption is a common practice among users because of the spread of smart devices.

The proposed services were widely accepted by users as an accessibility service, highly regarded as a social tool, and have the expected quality to encourage their deployment. The user evaluation process provided valuable information and feedback from participants.

Bearing in mind the results, it was decided to carry out further tests on delivery of AD content with special focus on smart devices (smart phones, tablets etc.). The aim is to gather end user feedback on usability and accessibility for such applications providing content adapted to the needs of people with sensory impairments.

Pear Tree project: Creating Audio Descriptions Across Europe

The Pear Tree Project, conducted as part of the DTV4All project, aimed at finding general characteristics of describing film narratives across 12 (including 10 European) languages in order to determine whether it is possible to develop common European AD standards, and whether it is possible to translate AD scripts across languages. The analysis shows that we cannot really generalise about film descriptions in these languages because the analysed texts manifested huge variations. The general assumption was that if there are no statistically significant differences among languages, the results might be interpreted as depicting characteristics common for all of them. However, statistically significant differences were found in 10 out of 12 examined aspects.

The differences across languages in the 10 remaining aspects are too huge to allow justified generalisations and, additionally, some opposing tendencies, or trends may be observed. For instance, the majority of descriptions did not include the word 'film' or its synonyms but, on the other hand, the majority of texts included allusions to film-viewer perspective. Irrespective of the findings obtained in the study, it was assumed that common European guidelines could be developed, provided they take into account linguistic and cultural differences as well as preferences of visually challenged audiences in the countries concerned. Along similar lines, we can assume that translation of AD scripts is possible if translated ADs comply with the prevailing norms and preferences in the respective languages.

From the general outcome regarding the possible translation of ADs across EU languages, in HBB4ALL we shall take a step forward and propose automatic translation of ADs performing also tests to check their acceptability.

2.2.ADLAB

The ADLAB[3] project started from the premise that the situation regarding access to audio-visual services on the part of the blind and visually impaired population in Europe should be improved. It aimed at addressing the needs of the blind and visually impaired communities in Europe through the provision of much higher quality Audio Description over a wide range of uses. More specifically, the ADLAB project aimed to (i) analyse current practices and national norms and guidelines (where they exist) to have a clear picture of the present situation and gain a clear understanding of what is common to all and what is still required in the provision of Audio Description; (ii) define a set of international standards and reliable guidelines for the industry and for all users; (iii) provide the material for the setting up of a Europe-wide network of Audio

[3] www.adlabproject.eu

Description courses in higher education; (iv) sensitise policy-makers regarding the importance of providing the blind community with access to audio-visual services.

ADLAB partners included universities in five European countries (Belgium, Italy, Poland, Portugal, Spain), and industry partners and service providers in three (Belgium, Germany and Italy). This set-up is already an example of hybridity primed for fruitful collaboration and, as one of the aims of the project was to extend cooperation between academia and the world of work in all European countries, the consortium could provide an initial model on how to create the mechanisms by which best practices can be shared and acted upon. The partners had also been chosen on the basis of a diversification principle: the language permutations involved are Catalan, Dutch, English, French, German, Italian, Polish, Portuguese and Spanish (to include both widely spoken and lesser-used languages from the Romance, Germanic and Slavonic families), and all forms of audio-visual translation are included (dubbing in Italy and Germany, monolingual subtitling in Portugal, bilingual subtitling in Belgium, voice-over in Poland, bilingual dubbing in Catalonia). The project has formulated standardised guidelines for audiodescribers, to be also used, in defining reliable material for higher education courses. Tests results will be published shortly [1].

Results from ADLAB will be fed into HBB4ALL when producing AD stimuli for the user tests, or for AD translation.

2.3.Speech and machine translation technologies

The ALST project, funded by the Spanish Ministry of Economía y Competitivdad (2013-2015, FFI2012-31024) and led by UAB addresses issues which can impact HBB4ALL. This is a summary of the research carried out in ALST in the field of machine translation (MT) and text to speech (speech synthesis) as applied to Audio Description (AD) and voice-over (VO).

Finished experiments on text to speech AD: 67 blind and partially sighted people were asked to listen to AD excerpts with four different voices (female natural, female synthetic, male natural, male synthetic) and answer a questionnaire which included an assessment of overall impression, accentuation, pronunciation, speech pauses, intonation, naturalness, pleasantness, listening effort, and acceptance. A post-questionnaire gathered additional information. In all items median scores were above 3.0 and in some items they were the same for natural and artificial voices. When asked about their preferences, 81% preferred a human voice in the AD, 1% preferred a synthetic voice, 3% said it depended on the audio-visual product, and 15% did not have any specific preferences as long as the artificial voice was natural enough and was not tiring. When explicitly asked about text to speech AD being an alternative solution to human voiced AD, 94% responded positively, the main reason being that it would increase the amount of AD (33%) because it would reduce costs and time (36%).

Finished preliminary tests on MT in AD (Spanish<>Catalan): this research focused on the MT of filmic AD as a means to increase accessibility in multilingual settings. The corpus contained 4,384 words divided in 442 sentences, and the engines selected were Google Translate and Apertium. The MT output was analysed using an error categorization model. 42.22% of the sentences produced by Google Translate contained at least one mistake, whilst the error rate for Apertium was 57.78%. The rest were acceptable units that required no postediting. At word level, Google contained mistakes in 5.56% of the words, whilst Apertium's percentage was 11%.

Ongoing experiments on MT in AD of fiction films (English>Catalan): this research aims at determining whether creating AD in Catalan ex novo requires more effort than translating it from English into Catalan and/or postediting its Catalan machine translated version. In order to select the engine to be used in the

experiment, a pretest was carried with 5 participants on five online free MT engines (Apertium, Bing, Lucy, Google and Yandex) (it must be stressed that funding was not available to train a specific engine). Subjective and objective measures were gathered (postediting time, HTER, PE necessity, PE difficulty, adequacy, fluency, ranking), showing that Google was the best MT engine for the purposes of the experiment. In the main experiment, 12 participants were asked to create AD, to translate an English AD and to postedit a machine translated AD of 3 three-minute-long clips from the film Closer. Quantitative (postediting time, keyboard effort) and qualitative (perceived cognitive effort) data were gathered and are currently being analysed.

Ongoing experiments on MT in VO of documentaries (English>Spanish): the aim of this research is to compare post-editing and human translation temporal, technical and cognitive efforts when translating wild-life documentary films to be voiced-over, and to assess and compare the quality of resulting translations and posteditings. Experiments have been done and data are currently processed. The next step will be to test text to speech voice-over in documentaries.

In HBB4ALL this existing research will be taken another step forward by testing full AD translation automation.

2.4. Clean Audio and speech intelligibility

Several small-scale projects and investigations regarding Clean Audio and speech intelligibility in TV programs have been conducted by IRT during the past years. Different parts of the broadcast chain were addressed. Many problems with speech intelligibility could be avoided if the capturing or the mixing of audio signals would be conducted properly. IRT (as the R&D institute of the German public broadcasters) contributed to the composition of a guidebook for the production of audio for TV [5]. This guidebook includes recommendations regarding the capturing of audio as well as the mixing of audio, and thus supports avoiding typical mistakes during the production steps. The guidelines were published in July 2014, are publically available and can be downloaded (German only).

The most current project regarding speech intelligibility at IRT deals with voice-over-voice (VoV) signals. VoV is typically used for non-native language content where a translation is mixed to the original signal. If the relation (in dB) of the translation track and the original signal is too small, the intelligibility of the native language translation is considered as insufficient. IRT conducted several investigations and listening tests to find a recommendation for the mixing level in such VoV situations.

The generation of Clean Audio signals for use in HBB4ALL will build directly on these investigations.

2.5. Automatic generation of Audio Description

TVC has an Automatic Audio Description Generator in production for the live broadcast workflow. Specifically in news programmes and exclusively when some person on screen speaks in a language that is neither Catalan nor Spanish. When the news piece is scheduled to be broadcast, if it has any text on screen in a foreign language (having the text and time code when it will be visible), the Audio Description Process is launched and generates the corresponding audio.

This system basically has two parts; a Text To Speech subprocess and a Publishing subprocess. The Text To Speech subprocess gets data from a manual translation used as text on screen and it generates as output as many audio files as texts on screen are defined.

Figure 2. News piece with a French person interviewed, the translation is shown as subtitle and is spoken at AD cannel.

The Publishing subprocess launches the reproduction of each audio file at the synchronised time at the Audio Description channel. Currently AD is being generated exclusively for broadcast and is not fully integrated in the IPTV service (HbbTV or any other IP platform). No single integrated audio track mix containing AD and the regular audio exists, the IPTV delivery chain does not take into account such an additional audio track, the publishing system does not take into account the necessary metadata and the playback systems do not expose this content at all. The aforementioned issues are planned to be addressed in HBB4ALL.

2.6.HBB-NEXT

In the EU-Project HBB-NEXT[4] inter-device and inter-media synchronization components have been developed allowing for synchronisation of multiple streams on a single device or of multiple streams presented on different devices. For both features different approaches have been implemented and evaluated. One approach is based on existing concepts for media synchronisation, namely to use PCR and PTS which come with MPEG Transport Streams, that has been extended to work with multiple streams. While it is a simple approach it also has some limitations on use cases, especially if content is pre-produced or is served from a third party. To enable such use cases a more sophisticated concept was developed that is based on adding content timelines.

For DVB broadcast services the so-called DVB timeline was chosen, that is delivered as separate information in elementary streams and which is used to synchronize the content from broadcast with on-demand IP video (and/or audio). This concept is quite flexible and can be implemented also with live streaming on the Internet or managed IPTV services using different system formats like ISOBMFF a.k.a. MP4 file format by providing a suitable timeline format. For inter-device synchronization, additionally timing information is constantly exchanged between devices to synchronise the internal clocks for each media player. On this basis the HBB-NEXT Settings Application showcase is a novel example to

[4] http://www.hbb-next.eu/

demonstrate the feasibility and necessity of synchronization services for both end-users as well as industry-like broadcasters. The Settings Application offers accessibility services via a unified interface, as depicted in Figure 3.

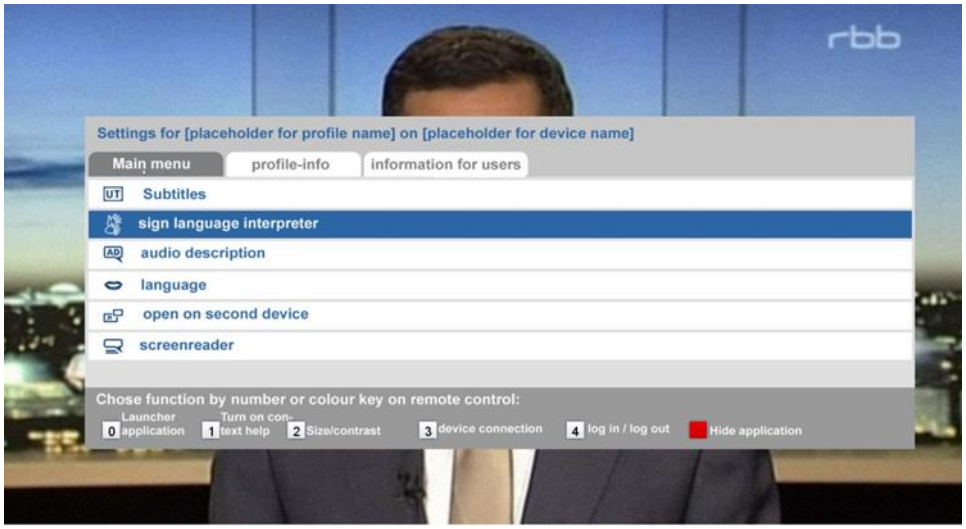

Figure 3. HBB-NEXT Settings Application showcase.

The service provides the following use-cases for media synchronization:

- Audio Description and IP subtitles and/or sign language video synchronized to broadcast video on TV.
- Inter-Device Synchronization of DVB-Video on TV to alternate IP audio-streams on connected device (e.g., Audio Description) and Synchronization of Subtitles to Video Playback on a connected device.

The DVB broadcast timeline approach to enable inter-media and inter-device synchronization in HbbTV requires technical amendments in all parts of the DVB broadcast-and-play-out chain, i.e. generating of an accurate timeline by the broadcaster, preserving the timeline in the delivery network, and finally using it to tightly synchronise multiple media streams on the end-user devices. Especially on the client side, a set of parameters need to be fine-tuned, since in parallel to the reception of the regular DVB streams, the broadcast timeline needs to be extracted and the play-out of "slave" digital media content needs to be buffered and controlled. These actions require a certain amount of processing time, which needs to be considered in the overall media play-out. In exchange for these efforts, absolute timing information is available on the client side. It even provides a useful basis for future developments in the area of multi-device synchronization: if the timeline is shared among end-devices, it may clearly bring forward frame-precise synchronized services for hybrid media on multiple devices.

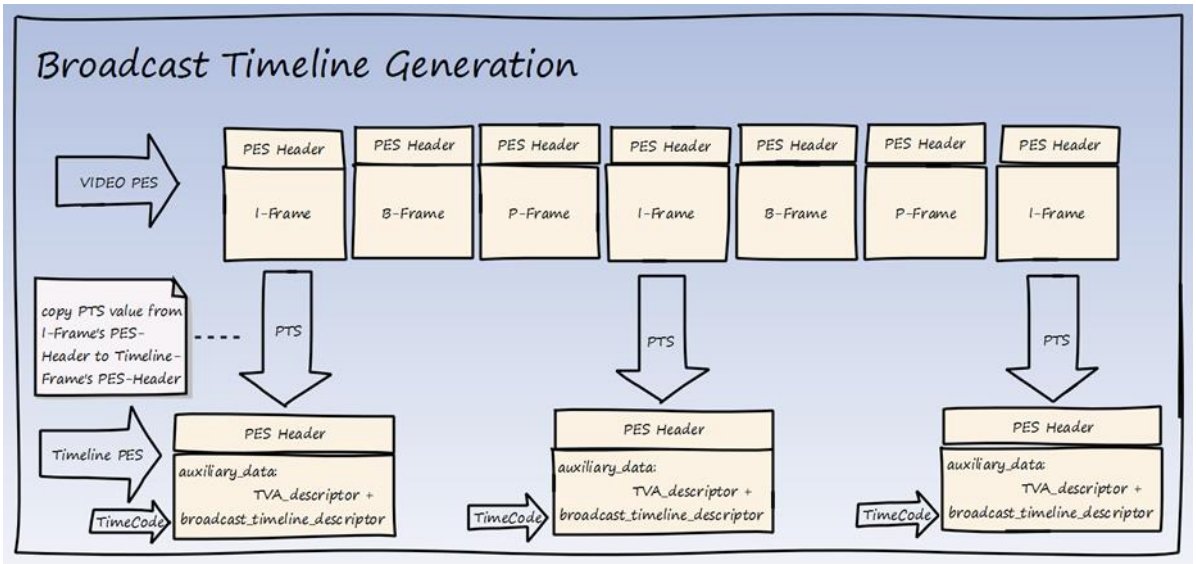

Figure 4. HBB-NEXT: timeline generation to support broadband bases synchronized services.

The ability to provide broadband-based synchronized services allows for providing services for smaller target user groups. Due to the signalization of additional content over IP, external companies can, for the first time, provide additional services to broadcast programs. This gives broadcasters the opportunity to outsource features (different language tracks, etc.) and increases the importance of synchronization in future business models in a connected TV environment.

In particular the deliverables of HBB-NEXT work package 4 are describing the HBB-NEXT technology to solve the problem of DVB to IP video synchronization in HbbTV applications and also the usability results[5].

Any Pilot-B activities with respect to HbbTV 2.0 will be based on the HBB-NEXT results (specifically the use of a timeline as a basis for synchronised services).

[5] http://hbb-next.eu/index.php/documents

3. Challenges, workflows and technologies

This chapter addresses the challenges that specifically broadcasters (as service provider) face, when they are looking at extending their service offer with additional audio services. The current workflows in audio production, as well as possible adaptations are addressed and the technologies that partners plan to use in Pilot-B for the implementation of the services are introduced.

3.1. Challenges for production and distribution of additional audio services

The production of additional audio services means several "high level challenges":

a) Increased number of audio assets

The most obvious challenge of producing additional audio services is the amount of audio content which has to be produced, delivered and provided. This influences the generation and the distribution of audio as well as the Graphical User Interface of an application providing options to viewers/listeners. For instance, it would be advantageous to use MPEG-DASH media playback capabilities as this avoids publishing multiple A/V files with one video track and different audio tracks. This requires several separate files for the audio to be generated, integrated into the workflow and finally published, including the relevant metadata to properly identify each additional audio.

b) Economic challenges

Providing additional audio services is also a question of finances; for broadcasters it involves considerations of effort and real benefits. The more steps in the workflow are automated, the easier and cheaper it is to integrate services into the broadcasters' workflow. Manual generation of additional services would mean increased and recurring personnel costs as opposed to a one-off investment in new technology.

c) Multiplatform delivery

To address multiple platforms (for example on-demand online media libraries and HbbTV) the additional audio services must be reproduced in various formats for each distribution mechanism; systems have to be implemented to render these different formats. Moreover, on top of different format requirements, playback capabilities are also different across devices. This means that there are significant differences in the level of service and functionality on each platform. Broadcasting delivery commonly already takes into account multiple audio options in DVB but that flexibility is rarely seen in multiplatform services.

d) Quality

The target group for Clean Audio are the hard-of-hearing, and every one of them has individual needs. Using the human ear to assess quality and suitability of audio samples for the majority of the target group is complex and challenging.

Depending on the type of input format (stereo or 5.1), the quality of the processed CA versions will be different. As the 5.1 signal already includes a separate channel for the speech, the result most likely will be better than the speech separation step for stereo signals (see section 6.1). Unfortunately, only for the high-quality productions (at least in Germany) multichannel content is available. Therefore one major

goal of the CA generation will be to find the best solution to separate speech from pre-mixed stereo signals.

Another challenge for the CA pilot will be the identification of the optimal mixing level of speech to the rest of the audio signal. The CA pilot basically aims at offering two different CA types (one including dynamic and frequency processing and one without) with several sub-versions which differ regarding the mixing level relation of speech to the rest (atmosphere, music and effects). It has to be investigated, which mixing levels shall be offered to the users.

Although the usage of the HBB4ALL CA service will provide better speech intelligibility for the target group, it might be possible that the listening experience suffers from the CA processing. Hence, it is important during the evaluation process to not only address the intelligibility but also the listening experience of the content.

Regarding Audio Description quality, the content generated by a professional AD operator does not present outstanding challenges other than ensuring the optimal mix between the AD and the regular soundtrack. On the other hand, AD content generated using TTS technology may have quality problems depending on the TTS implementation. In the case of TVC, quality issues have largely been solved for broadcast and therefore the most important challenge remains to ensure the same quality levels are applied to IPTV in general and in HbbTV in particular.

e) The “human factor” in the production department

Audio engineers are technically responsible for audio mixing, and it’s crucial to involve them from the beginning in the process of development of new services and technologies. This makes sense because of their know-how; without convincing the engineers, it would be very challenging to integrate this new approach into the production chain.

Directors are responsible for the dramaturgical realisation of audio mixing. It will be very challenging to convince them to offer mixes which significantly differ from the original. Developers should therefore try to provide mixes which are as close to the original source as possible and which take the original dramaturgy into consideration.

During the project, and especially in the pilot preparation phase, Work Package 4 partners plan to respond to these challenges.

3.2. Audio production workflows

3.2.1. Existing RBB Audio Workflow

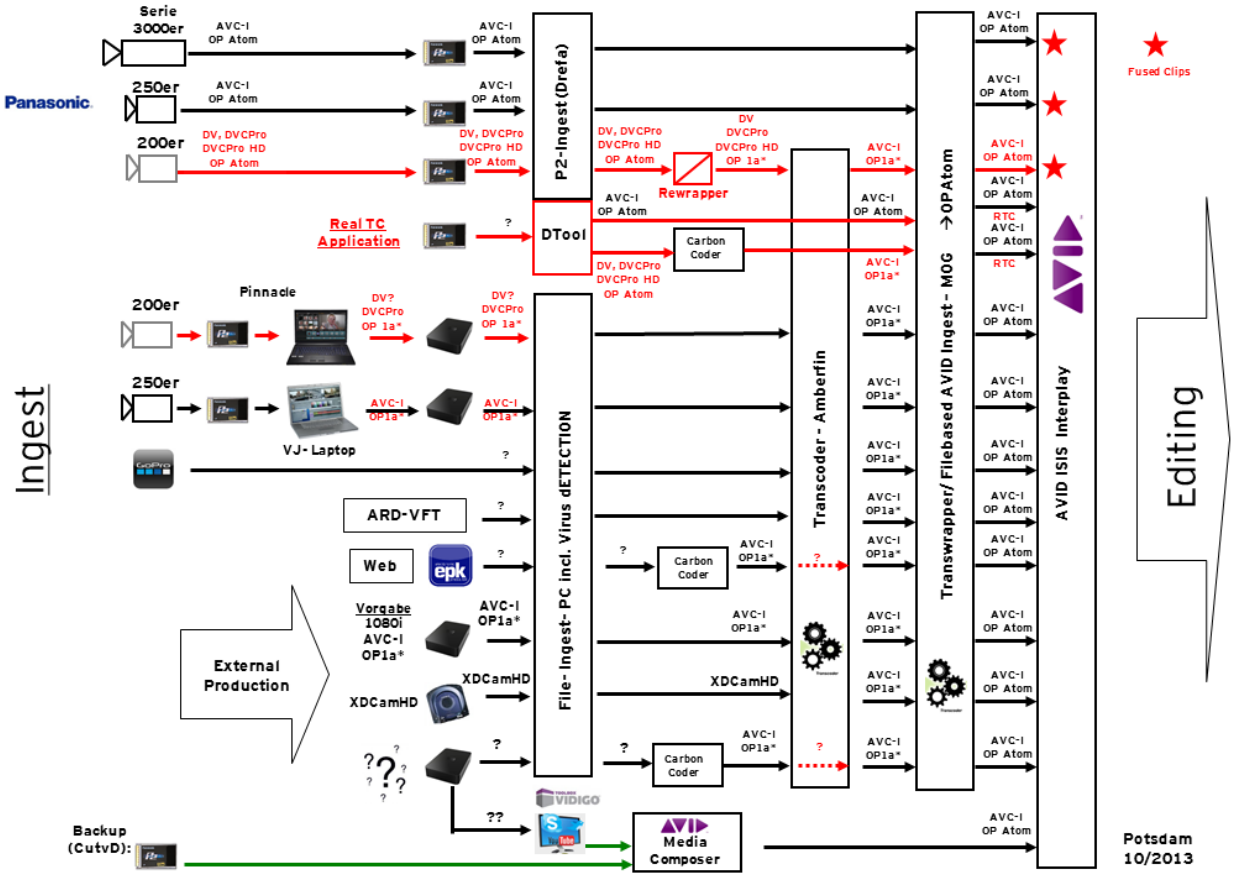

Figure 5. RBB ingest chart.

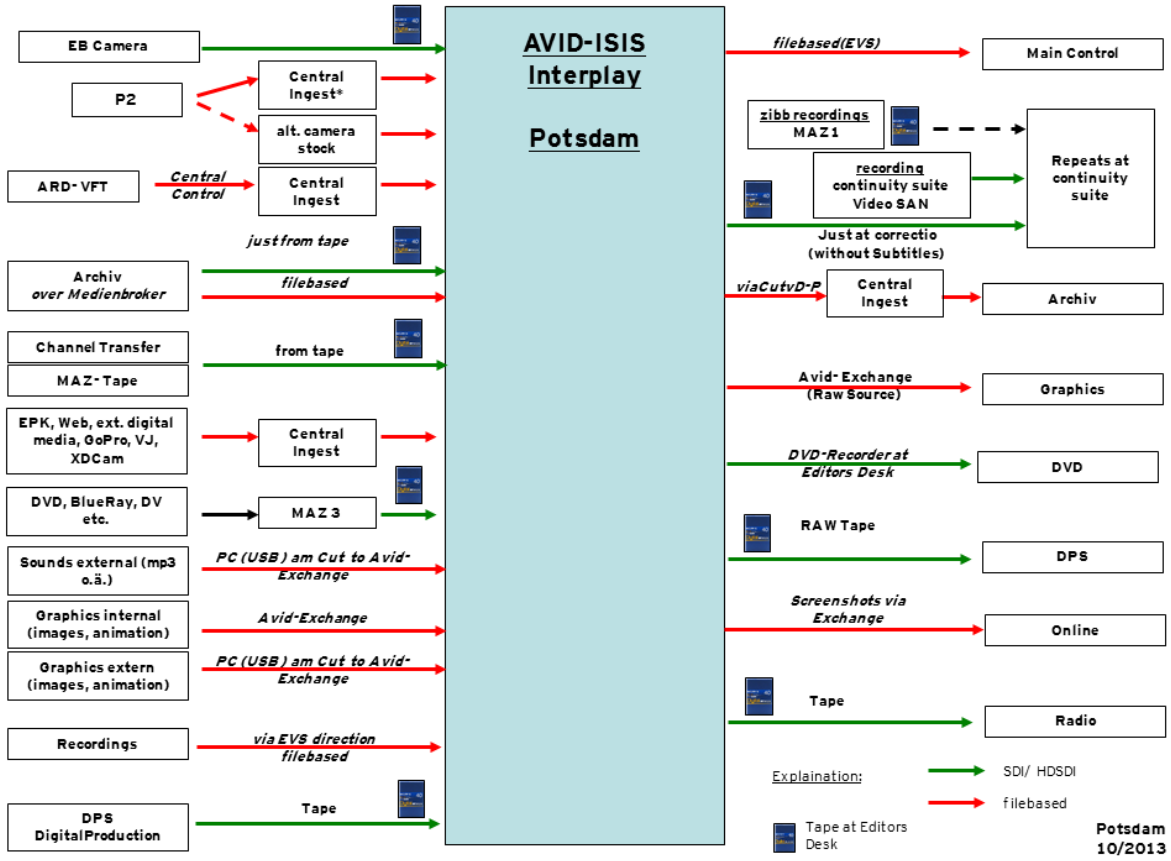

Figure 6. RBB production chart.

In the everyday production workflow of public broadcaster RBB, there are two common practices for audio mixing:

a) for pre-produced contributions, directors and editors jointly mix audio in the editing suite;
b) for live production, audio mixing is included in the overall RBB production process in the on-air control-room.

More complex and cost-intensive productions (e.g. fairy tales or crime series) are outsourced by RBB and produced by external production companies. Editorial decisions are of course made by the responsible RBB departments, but the technical workflow, including audio mixing, is fully outsourced.

The current workflow does not feature a CA production and processing. Given the diversity of sources and ingest workflows, enabling CA in multiple components would be very extensive.

3.2.2. Existing TVC Audio Workflow

The current TVC audio workflow is illustrated in **Figure 7**.

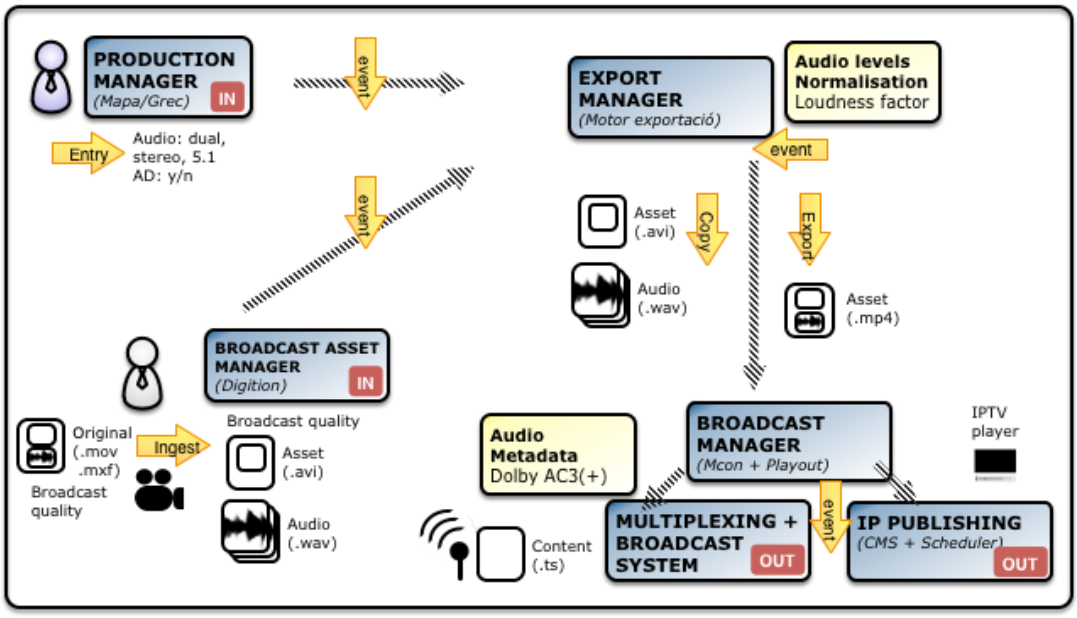

Figure 7. Existing audio workflow at TVC.

The assets are ingested through the Digition Broadcast Asset Manager and can have as many audio channels as necessary. All the audio channels can be stored at the Digition using the multichannel wav format.

For simplicity TVC has some predefined audio presets in order to support all the usual operational audio possibilities. These presets resolve the metadata audio definition but are not exclusive and moreover it is possible to define new presets or to include other audio channels with customized content if necessary.

Once the broadcasting of some asset is scheduled that it must be a production finished asset (asset ready for broadcasting), the corresponding event is released and the Audio Levels Normalisation process request is launched. This normalisation pursues avoiding the change of audio level perception between different assets at broadcasting time.

After normalisation the Export Manager will transcode the asset to the different broadcasting formats as needed.

3.2.3. TVC existing Audio Services

3.2.3.1. TV3 A la carta application

The on demand *TV3 A la carta* application, accessible through on-line webpage, offers many videos in both local and original language, but no other audio personalization functions (e.g. Audio Description or Clean Audio) are available.

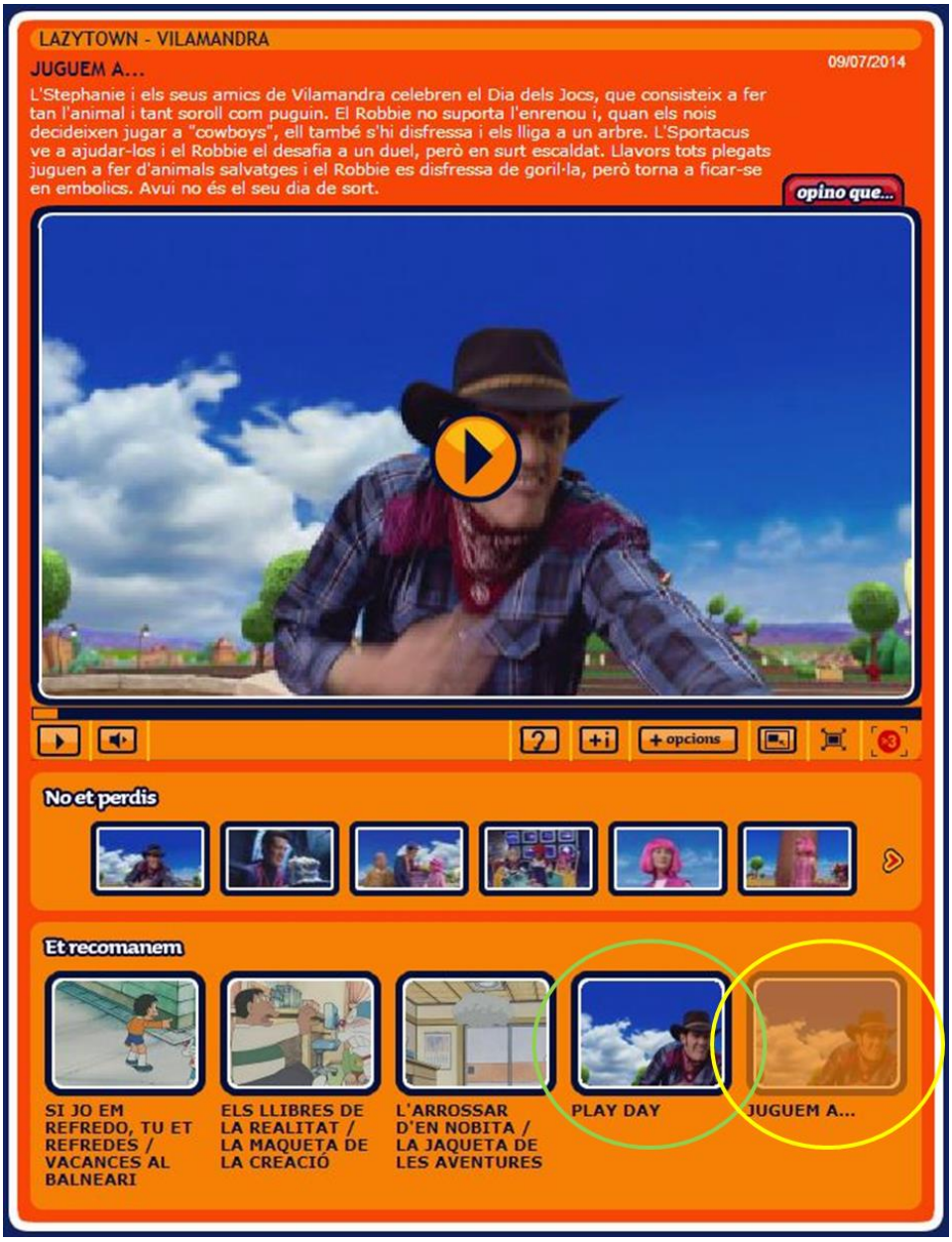

Figure 8. TV3 A la carta application.

On the example of children's channel *Super3* (see Figure 8) we do not have any direct option of switching the language of the video. Instead, for many TV shows there are available two separate videos of the same program, one in the local language (Catalan) and another one in its original version. As we can see in the illustration below, the video is streamed dubbed in the local language with the title of episode "Juguem a…" (see yellow circle). Next to it, there is a video with the same key frame and the same title but in English - "Play Day" (see green circle), which is the same video but in its original version.

This kind of navigation needed for switching the language might be quite uncomfortable for the user and not visible enough: whilst watching the video in Catalan there is no information that the same video is available in the original language. The video in a different language is merely displayed as one of the recommended contents. Also it occupies twice the GUI space than just putting one video with a changeable audio stream.

3.2.3.2. Broadcast TV services

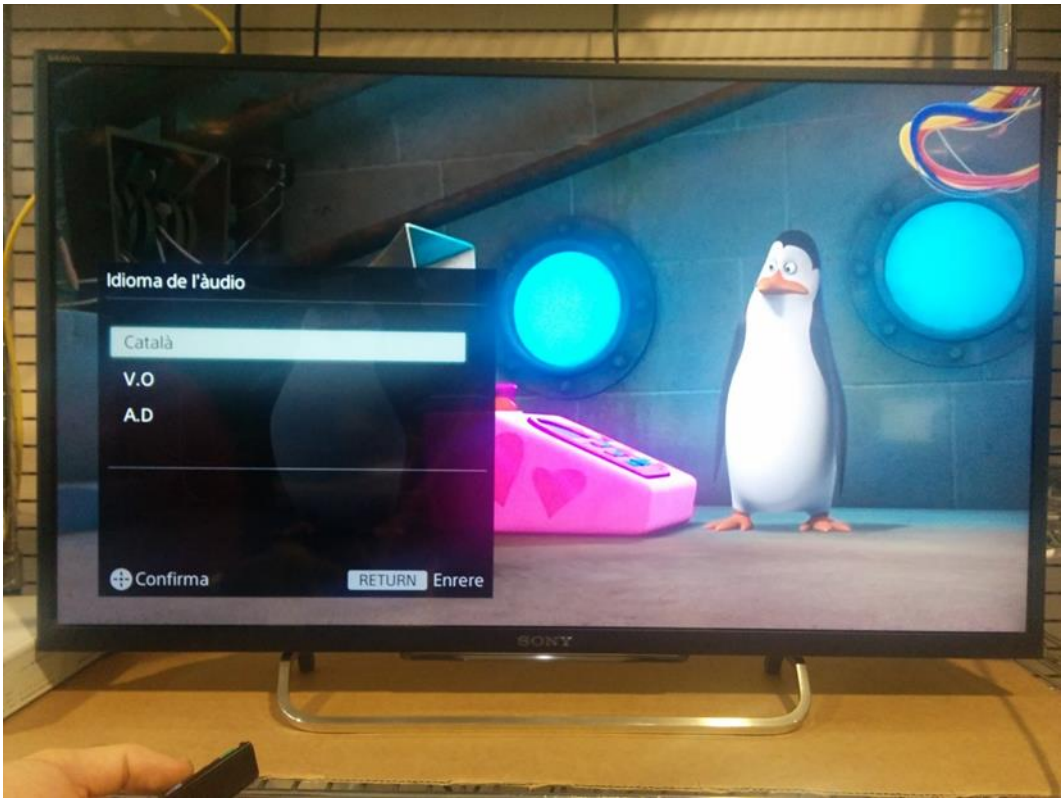

Figure 9. Audio settings from the TV device.

While watching one of the TVC channels through Digital Terrestrial Television (DTT) there are available more audio settings than via HbbTV application. For all programs in foreign language there is an option of enabling the original voice.

The button to access the audio settings is situated directly on the remote control or after clicking "Audio language" from the main menu (depending on the type of TV device). After calling to action, the user can choose from following types of audio (see Figure 9):

- Catalan
- V.O (Voice Original)
- A.D (Audio Description)

The Audio Description is available only in the local language, which is Catalan. After clicking one of the buttons, the selected audio track is played seamlessly.

3.2.4. Workflow changes required to support additional audio services

Identifying exactly where workflow changes are required depends on the specific technology. For existing workflows at RBB, it seems reasonable to generate the Clean Audio content on the basis of the already-mixed content. It would be very complicated to set this up beforehand in the workflow chain, as technology and working steps would have to be changed for each editing suite and in each on-air control-room. This would imply significant expense. Therefore the integration of new technology after the mixing process seems the better option. However, the technological approach/concept should be concretely established first and only then, when the project is running, it can be evaluated where exactly the changes in the workflow should take place.

At TVC, workflow changes required are found in two main areas. Firstly, the relevant content needs to be generated appropriately for inclusion in the digital distribution delivery chain. In the case of original audio soundtrack, the original audio content needs to be identified, tagged and transcoded in the appropriate end format and codec (usually AAC+). In the case of Audio Description content, the existing workflow generates the content only for broadcast (in the form of separate audio files for each AD sequence) which means content needs to be aggregated but also identified, tagged and transcoded for digital delivery. Finally, content prepared for HbbTV delivery in the workflow needs to be published appropriately for MPEG-DASH, which includes manifest generation, asset transcoding and metadata preparation (which assets have multiple audio and which do not), specifically taking into account if MPEG-DASH is used to select playback of individual audio tracks or A/V are merged into different assets.

3.3.Distribution/media delivery platforms

The video production includes multiple stages. Normally videos are produced with a focus on OnAir broadcast. Preparation for online services such as Internet or HbbTV is often a downstream process. At RBB there is a variety of online production workflows, see Figure 10; online production starts with range of sources from Digital-Beta to DVB-T recordings. A very small amount of preparation is done with sources taken directly from original video. Any Clean Audio tracks additionally produced would be lost during these common workflows and processes.

The final processing for online video is usually transcoding/packaging to satisfy the needs of different technical distributions. The videos are transcoded into different video-formats, and different qualities. The last step is segmenting and packaging for adaptive streaming like HLS or MPEG-DASH. This is usually based on services of the Content Distribution Networks (currently most ARD broadcasters use Akamai Universal Streaming Service). It must be determined whether additional Clean Audio tracks can be processed by these services. A Clean Audio generator would be ideally placed beside the video transcoding components.

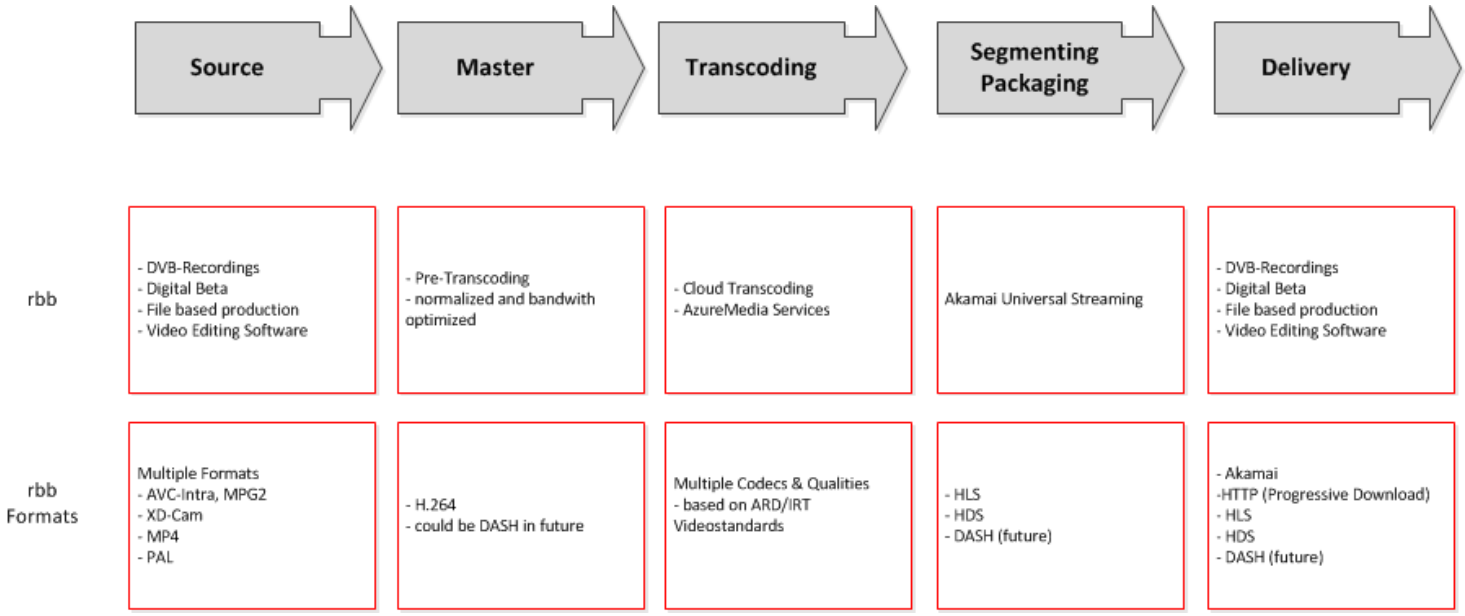

Figure 10. RBB workflow for online video distribution.

In the case of media delivery and distribution at TVC, two main areas are identified: firstly, publishing the necessary metadata using HTTP and conventional CDN caching, to allow the HbbTV application to display the content appropriately, clearly indicating which audio options are available. This is done in the appropriate file format such as JSON or XML, easily readable by the JavasScript logic. These content files, being plain HTTP objects, are easily cached by any modern CDN. Cache invalidation is usually done 'passively', that is, by generating new asset identifier URLs, which are read by the CDN as new content.

Secondly and critically, MPEG-DASH content needs to be delivered through the CDN for optimum performance. Though all content delivery is strictly HTTP, the CDN needs to be aware or at least tuned in accordance of the specific nature of chunk-based media delivery as it is the case of MPEG-DASH. This includes but is not limited to the following aspects:

- Delay to start playback due to client buffering of one or more chunks
- Delays in A/V stream selection by the player
- Different caching strategies and tuning for manifests and chunks
- Cache performance worst-case scenario degradation as old chunks are obsolete very quickly in live MPEG-DASH distribution
- Tuning of manifest caching in live distribution as opposed to aggressive caching in on-demand manifests (and chunks)

Not unlike plain live video distribution in IPTV in general, live generation and distribution presents its own challenges (such as server workload, complex configuration, content quality, excessive buffering delays, etc.), which also apply to MPEG-DASH generation. This is further compounded by generating and distributing the adaptive quality capabilities of MPEG-DASH. Due to these constraints, pure on-demand distribution is usually the first environment to test any new services (such as it is the case for the planned service pilots, see section 4). Moreover, a "live" scenario most likely cannot be supported during the project's lifetime (see section 3.5) so any development, testing and deployment in Pilot-B is targeted for on-demand distribution.

3.3.1. Changes required to support additional audio services in HbbTV 1.0, 1.5 and 2.0.

Pilot-B focusses on the provisioning of multiple audio services in addition to a TV program, and targets at using the installed base of HbbTV devices in the market. For Clean Audio it is foreseen to provide multiple varieties of the CA audio signal within the same service; Audio Description and additional language tracks will constitute further audio streams related to a TV program. With respect to HbbTV following technological options are considered within the scope of HBB4ALL for the delivery of multiple audio streams with a TV program:

HbbTV 1.0 devices:

- A live use case can be supported by delivering multiple audio elementary streams via DVB broadcast. However, tagging the appropriate stream to allow selecting it from an HbbTV application can only be realised in a non-standardised way. In the HbbTV 1.0 specification, audio components can only be identified by language, encoding and whether they are signalled as Audio Description.
- An on demand use case, where the content is completely delivered via IP (broadband), can be supported but requires for each audio version that is part of the service, that the content is prepared and stored in a separate ISOBMFF (MP4 container) file. Each file contains the video multiplexed with one version of the audio. Switching between streams at playtime is possible (the player switches from one A/V file to another). The user will notice a short interruption in the playback due to buffering the new content and seeking (jumping) to the position where the old content stopped. This use case is supported by all HbbTV devices in the market.
- The synchronisation of an audio stream delivered via IP with the video transmitted via DVB broadcast is not possible with off the shelf HbbTV1.0/1.5 devices.

HbbTV 1.5 devices:

- The live use case via DVB broadcast is also valid for these devices. On top of the 1.0 functionality, with HbbTV 1.5 the PID (packet identifier of the elementary stream) and a component tag if signalled with a stream identifier descriptor can be used to distinct between different audio components in a service. This is useful to provide the user some meaningful information about each single stream before he makes his selection.
- Both live and on demand use cases, completely running via IP (broadband), can be supported by means of MPEG-DASH (see section 3.5.1). The various audio streams as well as the video content must be prepared on the provisioning side. The management of the resources and various audio / video components belonging to them, must be handled and presented transparently to the end user. As opposed to the 1.0 case, switching between different audio components at play time now merely involves the audio stream. The DASH profile of HbbTV 1.5 allows for 16 adaptation sets, i.e. audio and video components. This means that if there is one video component, one DASH presentation can include up to 15 audio components.

HbbTV 2.0 devices:

- Based on the Multistream Synchronisation feature (see section 3.5.2), the live use case via DVB broadcast can be enhanced by synchronisation of additional audio streams delivered via IP (broadband). It is required that such additional streams are preproduced, to allow the streams

delivered via broadband to be exactly timed with the broadcast signal without the need to buffer it on the user end device.

- Multiple audio decoders are optional; thus client-side re-mixing of multiple audios could also only be optional.

With a single HbbTV application, that is aware of the different HbbTV implementations (1.0/1.5/2.0), it is possible to cover the various HbbTV versions; nevertheless, care must be taken by service providers (broadcasters) to provide the different varieties of content to support the different versions, specifically if they want to support all HbbTV terminals currently available in the market and also want to use the features of the new versions.

3.3.1.1. HbbTV-related audio distribution - RBB view

To support HbbTV 1.5 the according Clean Audio content will have to be prepared in the according audio formats and published for use with MPEG-DASH (see also section 3.2.4). The correct setup for MPEG-DASH publishing (in combination with an appropriate HbbTV testing application) are currently being prototyped by IRT.

In order to develop multi-stream synchronization in HbbTV 2.0 it is necessary to integrate a timeline in the DVB stream at the playout centre of RBB. Any changes in the DVB signalling process must be prepared early and carefully. RBB and IRT are currently planning making a first technical test phase which will be conducted in 2016 at the earliest. Before the test phase in the real world environment, IRT and RBB will develop technical proof of concept showcases as preparation for the effective integration at the playout centre.

3.3.1.2. HbbTV-related audio distribution - TVC view

To take advantage of HbbTV 1.5, the changes outlined in section 3.2.4 need to be implemented to effectively support MPEG-DASH. Additionally, TVC plans to take advantage from the experience and the workflow changes from HBB4ALL in order to offer the same services at mobile, tablet and desktop web platforms, using HTTP progressive download distribution wherever feasible.

To take advantage of HbbTV 2.0, no further changes are expected due to the new capabilities, for instance, any extra audio services are already present in DVB-T and there is no need to use multistream synchronisation Additional content in the project will be made available mainly in the aforementioned VoD services, which do not require HbbTV 2.0 features.

3.4.End User Requirements (GUI)

In general, the hard of hearing target group is very heterogeneous - there are many subtle gradations of ability and need, and also many very different technical solutions. Thus there is no one-size-fits-all solution; in this project we should focus upon a variety of possible solutions.

RBB focuses on the living room television scenario. The goal is to develop technical solutions for the everyday life TV consumption. This means that the CA feature must be integrated very smoothly in the already existing TV environment. For the users, it should be very obvious which features are offered; there should be easy access to features, and features should be relevant to the targeted user. The graphical user interface should be very simple and not particularly ‘playful’ – the service is a function and not a game. In

order to address their needs, users will be involved in GUI development. In the case of TVC, the same approach applies, the intended service is not an extra feature to be 'enjoyed' but rather a required function of the IPTV ecosystem. In that view, the intended interface is not one of ease of discovery but rather of ease of access of a service known to users (especially in the target groups).

The remainder of this section focuses on a formal process carried out by TVC, to obtain requirements regarding the GUI of an appropriate application for the operational phase. Further requirements e.g. with respect to the audio content generation have been and will be obtained from the respective preliminary user tests that are carried out for Pilot-B. For detailed information on these requirements please refer to sections 7.1.1 (Clean Audio), 7.1.2 (Audio Description) and 7.1.3 (Other Languages and Language Learning) respectively.

3.4.1. TVC User Stories

In the audio service pilot (Pilot-B) we came up with various user stories through the process of brainstorming. We are aware that not all of them are doable at this point, that is why a previous requirement investigation is needed. All user stories for audio service are listed in Table 2 in order of priority:

US number	US description
TVC14	As an end-user I would like to have Audio Description available when I am not looking at television set or I am in the car or metro in order to enjoy the content at any moment
TVC15	As an end-user I would like to have access to a dictionary or lexicon while listening to the original voice of a program to learn difficult words
TVC16	As an end-user I would like to be able to clearly distinguish the Audio Description from the dialogue to understand better what is happening
TVC17	As an end-user I would like to hear better the dialogues that are in original voice and not in my native tongue to understand all
TVC18	As an end-user I would like to nearly eliminate other sounds only when there is a dialogue to better understand what they say
TVC19	As an end-user I would like to regulate the level of intelligibility added to the dialogue to fit my needs
TVC20	As an end-user I would like to have the original voice available at any moment and know which contents have it to learn languages
TVC21	As an end-user I would like to have a list of the contents that are in a specific language of original voice to facilitate the access
TVC22	As an end-user I would like to regulate the level of detail of the Audio Description to adapt it to the level of desired degree of attention
TVC23	As an end-user I would like to be able to choose default audio language in order to not having to re-adjust my preferences every time I enter the application
TVC24	As an end-user I would like to have an option of adjusting the volume of dialogues and background sounds separately

Table 2. User Stories (US) of audio pilot.

Three user stories most desired are described with more detail below.

The first user story of audio pilot (TVC14) concerns a basic need of the Audio Description service. Apart from users with vision problems, people that in a particular moment cannot look at the TV set or are outside home, for example driving car or using metro, are also interested in Audio Description, so as to enjoy the content at any moment.

In the second user story of this pilot (TVC15) the user shows the need for personalized audio settings to facilitate the language learning. In the case that the program is streamed in original language and there appear some words that might be difficult for the listener, there will be provided a dictionary and/or an explanation of the particular word or expression to foster language learning.

Finally, the user story TVC16 highlights another basic need, which is the clear understanding of the audio. When the Audio Description function is activated, the user would like to be able to distinguish it easily from the standard audio for a seamless experience. For that, the voices of dialogues and the one of narrator will be clearly distinctive in this scenario.

Not all user stories are feasible and that is why we have created only storyboards of those that we think can be considered at the moment. The user stories: TVC16, TVC17, TVC23 and TVC24 have to be initially investigated in terms of viability. On the other hand, user stories TVC15, TVC18, TVC19 and TVC22 are not feasible at the moment and need to be reconsidered for the future development.

3.4.2. TVC Storyboards

TVC14: As an end-user I would like to have Audio Description available when I am not looking at television set or I am in the car or metro in order to enjoy the content at any moment.

Figure 11. Storyboard to User Story TVC14.

TVC20: As an end-user I would like to have the original voice available at any moment and know which contents have it to learn languages

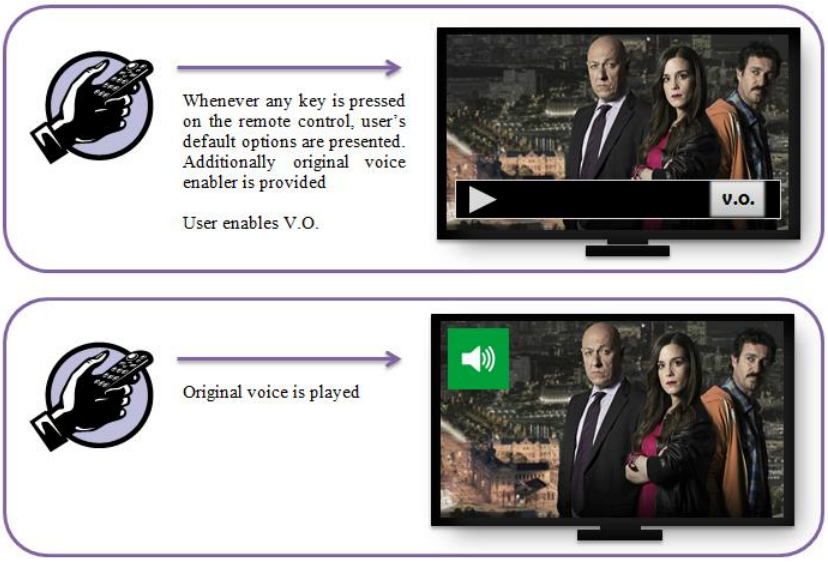

Figure 12. Storyboard to User Story TVC20.

TVC21: As an end-user I would like to have a list of available content in each specific language to facilitate access

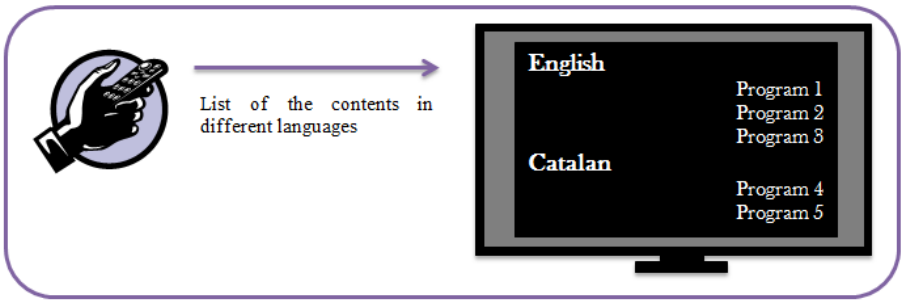

Figure 13. Storyboard to User Story TVC21.

3.4.3. TVC Mock-ups

For the Spanish services in Pilot-B we have implemented the same methodology as in the Pilot-A (subtitling services), i.e. in the first place we have created various versions of paper mock-ups to visualize contrasting alternatives of the service interface. Subsequently, we have carried out an internal workshop with the aim of deliberate possible options and select one of them for the future process of service development.

We have decided to start with creating paper mock-ups as they permit to carry out smooth modifications by cutting off the pieces and sticking buttons, which helps in apprehend all possible options at the moment and even construct new ones during the workshop. On the other hand, it is much more challenging to modify any element in the digital mock-up and additionally it requires a technical knowledge. That is why we will not create any digital prototypes as the interface is simple enough, so that the next step will be directly creating a real service pilot.

3.4.3.1. Paper prototypes

As it was explained before, we will start from creating basic paper versions of mock-ups. We have based its elaboration on user's requirements described in sections 3.4.1 and 3.4.2. For creating them, we have used paper sheets, pen and marker and also buttons and menu displays created through the computer, printed and cut in pieces for a flexible representation. Thanks to this ease of manipulation we have developed four different paper mock-ups, which are presented and described in the following sections.

The scope of the new feature is only linked with the video streaming so in the creation of paper mock-ups we used only one display, the one of the video in the full screen mode, as the rest of the TV3 *A la carta* application will remain the same.

To arrive to the mock-up display, which is the video streamed in the full screen mode, the following path has to be completed:

"User starts watching DTT channel of TVC, when a pop-up window shows up as a call to action to enter the HbbTV application TV3 *A la carta*. When the user calls to action by pressing the red button, he accesses the initial page of the application. In the overhead menu bar there are buttons like: Home, Programs, Last Week, Most Viewed, TV3 Live and a search engine. We can choose to see content between highlighted videos and highlighted programs or go directly to the "Programs" section and choose content from an alphabetic list. When the user decides to see a specific content, the video is displayed straightway in the full screen mode. By pressing any button on the remote control, an overlaid control bar appears with the common buttons like: pausing, minimizing or scrolling the video. Moreover, specific additional buttons regarding the pilot services will be shown and explained for each case in the following sections."

So far, we have performed two workshop sessions for the process of creating paper prototypes. In the first of them we have conceptualized possible alternatives and functions to demarcate the scope, which resulted in creating four alternatives of the paper mock-up described in the following sections. After that, we have carried out another workshop in which one of them has been chosen.

Paper Prototype Version 1

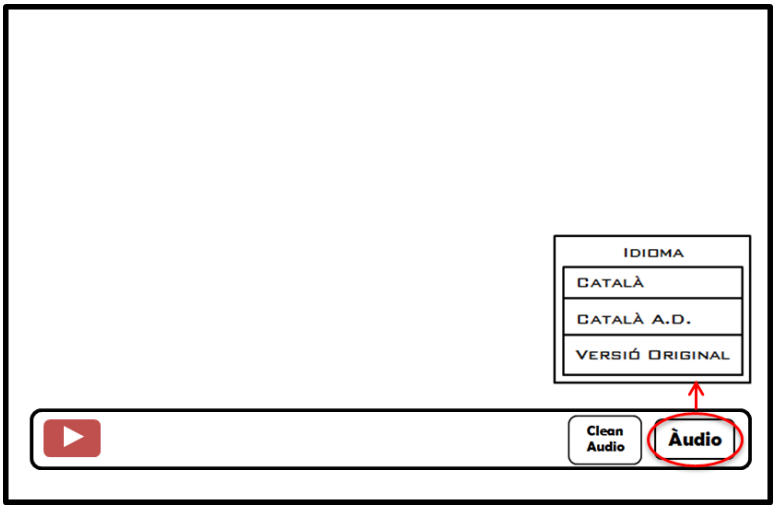

Figure 14. Paper Prototype Version 1.

The first version of the paper mock-up shows a scenario in which there are two buttons situated in the control bar: “Clean Audio” and “Audios”. First of them activates Clean Audio, which is a dialogue without any other sounds such as music or noises; all that with the objective of delivering an intelligible content. It can be either activated or deactivated, i.e. no more adjustments are provided for this function. The second button includes audio settings, such as language choice between Catalan, Catalan with Audio Description and original voice. The mock-up and its flowchart are presented in Figure 14.

Paper Prototype Version 2

In the second paper mock-up we have reduced the number of buttons from the control bar including all the functions in a single button “Audios”. After clicking the button a window with principal menu pops-up and we can configure the audio depending if we would like to hear the Audio Description (only for Catalan language) and additionally if we want the audio to be cleaned from sounds other than dialogues. It also allows to choose the audio language amongst Catalan and original voice.

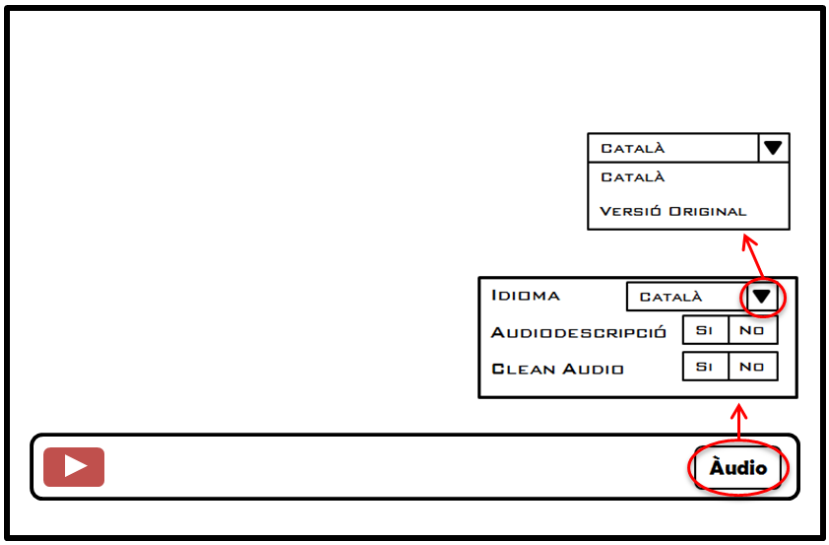

Figure 15. Paper Prototype Version 2.

Paper Prototype Version 3

In the next version of the mock-up we have presented all possible audio options in the control bar, i.e. “Clean Audio”, “Audio Description” and “Audios”. This time we dispose of more detailed adjustments in terms of cleaning the audio, that is to say, we can choose the degree of importance of dialogues and other audios by choosing a suitable division between those two elements. The “Audio Description” remains a function that can be either activated or deactivated. Finally, the button “Audios” changes its design adding to it a settings icon, implying that we can adjust here our audio preferences in terms different than two previous functions. In this case only language choice can be made between Catalan and original voice.

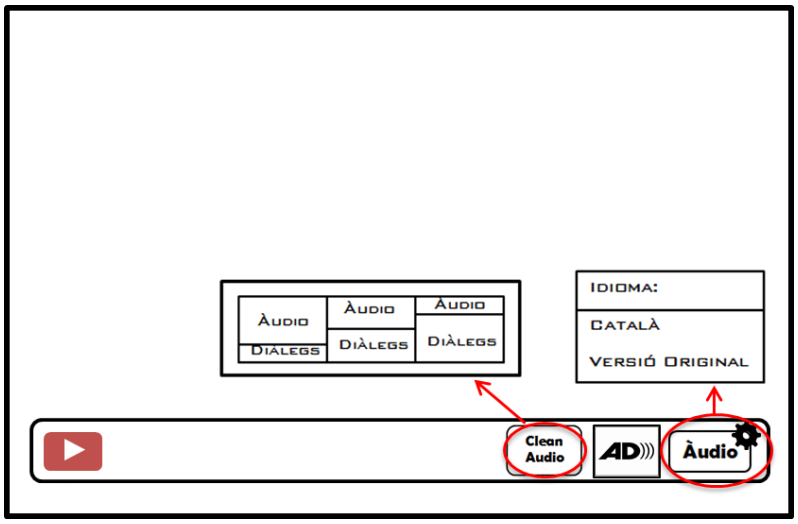

Figure 16. Paper Prototype Version 3.

Paper Prototype Version 4

Finally, the fourth and the last version of the paper prototype of the audio pilot has changed its interface situating only two buttons in the control bar: "Audio Description" and "Audio settings". The user has a direct access to activate/deactivate Audio Description function from the main control bar and additionally has another button where he can configure other audio variables. After clicking audio settings the user can elect language and also enable Clean Audio by choosing one of the following options: Catalan, Catalan Clean Audio, Original Voice an Original Voice Clean Audio.

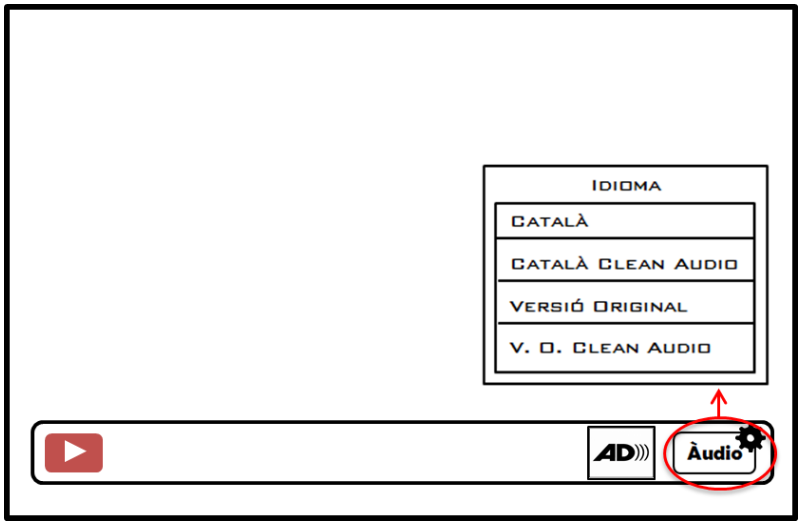

Figure 17. Paper Prototype Version 4.

Wrap-up

After creating those four versions, we have performed a second workshop session with the objective of presenting all options and choose one of them for the next step, which is the creation of the future pilot. Thus, the mock-up chosen in this workshop was the "Paper Prototype Version 4".

We believe that this alternative is the most suitable as it consists of a simple and intelligible interface. We think that the Audio Description function should have a direct access from the menu bar as it is destined primordially for blind users and because of this disability, the number of steps to active/deactivate should be minimal for an easy finding or even there should be a button associated to the Audio Description directly on the remote control. The rest of the functions like Clean Audio and language choice should be integrated jointly in the audio settings.

3.5. Technologies

This section introduces the technologies defined in HbbTV 1.5 and 2.0 that can potentially be used for the services in Pilot-B. These are 1) MPEG-DASH to enable adaptive live streaming over the Internet and 2) multi-stream synchronisation to deliver audio signals as separate streams via Internet in addition to a broadcast service. Also, an outlook is given on the pilot-specific technologies, which will be implemented as components for the realisation of the audio services.

It is noted here, that due to a delay in the publication of HbbTV 2.0 and more importantly (resulting from this delay) the foreseen unavailability of end user devices supporting HbbTV 2.0 in the project's lifetime, the pilots will be based on the installed base of HbbTV 1.0/1.1/1.5 devices. Services cannot yet make use of specific HbbTV 2.0 features. Specifically for Pilot-B this means, that the "Multi-stream synchronisation" feature will be missing, and thus a scenario with a live TV program (via DVB) enhanced by synchronisation of additional audio stream(s) delivered via IP (broadband) cannot be realised.

3.5.1. MPEG-DASH

As the name indicates, DASH = dynamic adaptive streaming over HTTP, MPEG-DASH makes use of the HTTP delivery protocol and provides a mechanism for seamless adaptation of the media quality to the available resources like the network bandwidth. MPEG-DASH defines different base profiles for on-demand and live use cases and for MPEG TS and MPEG ISOBMFF based media. From those profiles several standardization bodies have chosen the ISOBMFF live profile of MPEG-DASH including the HbbTV group for version 1.5 (ETSI TS 102 796 1.2.1).

Support of MPEG-DASH by the industry is increasing, encoders, CDNs and TV terminals are already available in the market. First services are provided in HbbTV V1.5 markets like France. In Germany measurements have shown that already 10% of the HbbTV devices in the market conform to HbbTV1.5 and hence support MPEG-DASH.

In the HbbTV V1.5 profile of DASH only audio, video streams but no subtitles are supported[6]. The specification also defines the interfaces for DRM but does not mandate a particular system. This is intentionally left open by HbbTV to the markets where it is deployed.

[6] Theoretically it would be possible to relate a maximum of 15 audio streams to one video stream (HbbTV1.5 profile of MPEG-DASH allows a total of 16 so-called adaptation sets with maximum of 16 representations each).

With HbbTV version 2.0 the DASH profile will be updated, it is backwards compatible to 1.5, but includes a number of changes for optimisation and it supports now also subtitles based on TTML and an event mechanism like stream events for broadcast. HbbTV refers to the DVB DASH specification that is published as a blue book on the DVB website.

3.5.2. Multi-stream and inter-device synchronisation

HbbTV version 2.0 specifies a new feature that allows to synchronise multiple streams on a single device or even on different devices in the home network, the two sub features are accordingly called multi-stream and inter-device synchronisation.

The two features are related and share the architecture and technical solutions. Many of the concepts and the architecture can be found in the DVB specification for Companion Screens and Streams (DVB CSS) that also has been published as blue book on the DVB website. While the use case for DVB is mainly inter-device synchronisation of a mobile device with a (DVB) TV, HbbTV has adapted the specification also for multi-stream synchronisation.

The following lists the details of the feature and their dependencies on other features:

- Synchronisation of one video, audio and subtitle component – where each of the components can either come from a broadcast service or as a pre-produced broadband stream – is mandatory. The limitation to pre-produced content comes from the optionality to buffer the broadcast service on the terminal.
- Synchronisation of a broadcast service with a broadband stream where both contain video is only mandatory if the TV receiver has sufficient resources to render multiple videos.
- Synchronisation of a broadcast service with a live broadband stream where the broadband stream is not available in time to be synchronised with the broadcast is only mandatory if the TV receiver supports the additional buffer for media synchronisation. The size of the buffer is calculated to work with a delay of a typical MPEG-DASH distribution via CDN and a typical maximum bitrate of an HD broadcast service that needs to be cached.

In summary the scenario for this work package where the additional audio services are delivered as an extra stream is mandatory for every HbbTV 2.0 terminal. Only one audio can be rendered simultaneously, which means that client-side re-mixing of multiple audios could only be optional. As mentioned above, audio content that is sent via broadband must be pre-produced.

3.5.3. Pilot-specific technologies

The service components that are specifically adapted to enable the audio services, also constitute relevant technologies with respect to Pilot-B. The components for the automatic generation of Clean Audio and Audio Description base on existing technology and will be adapted and improved for deployment in the pilots and subsequently integrated in the respective audio production workflows. Here specifically the issues of in- and output signals as well as (file) formats must be considered. Also mechanisms to manage multiple audio assets at the service provisioning side and to allow selection of the desired audio track by an end user will be included.

For detailed information on the Pilot-B service components, please refer to chapter 6.

4. Service pilots

Within HBB4ALL, the pilots have been defined as separate workpackages within the project structure. Each workpackage/pilot has foreseen several service pilots in an operational phase, in which for a specific pilot objective all service features will be included. Service pilots are foreseen to match each of the Pilot-B objectives:

- Clean Audio
- Audio Description
- Other languages and language learning

In each of the pilot regions (Germany – Berlin-Brandenburg, Spain – Catalonia) regional sub-pilots are planned (see Table 1. Overview of Pilot-B objectives and related sub-pilots). Most likely the sub-pilots in Spain will be carried out as a combined service pilot offering the functionality of all alternative audio services to the end user.

This chapter overviews the currently planned service sub-pilots in each of the pilot regions.

4.1. Clean Audio – Sub-Pilot in Germany (Berlin-Brandenburg)

RBB/IRT will implement a Clean Audio (CA) sub-pilot in Germany. At the time, the plan is to start the sub-pilot in Germany, in the region Berlin-Brandenburg, in autumn 2015 for several months. Ideally, the content will be made available for several months in the RBB online media library. A test panel from 10 to 20 users from the target group will evaluate the service at least for a couple of weeks.

4.1.1. Status of work at M12

During the first project year the technical approach has been adapted regarding the workflow requirements of the broadcasters and also based on the outcome of the first test with the focus group and subsequent lab tests in July 2014. The focus is now on using already mixed audio files as input signals, like 5.1 and stereo tracks. The status and enhancements with respect to the Clean Audio generator are described in detail in section 6.1.

Also approaches have been identified how the Clean Audio features can be integrated in the RBB workflow and how the features could be offered to the end users. Also the communication with the interest associations was established very well, to have a lively exchange and a productive collaboration in the coming months. This will help in understanding and respect the needs of the target group and will be also the basis for the user acquisition.

4.1.2. Timeframe

The timing is dependent upon developments during the next months and especially on the outcome of the lab test early 2015. Afterwards the further timing will then be decided upon jointly with the production and programming department at RBB.

The plan is to start the sub-pilot phase in autumn 2015 at least for a couple of months.

4.1.3. Test Users

Whether the sub-pilot will be an open pilot for all RBB viewers or a hidden test for a dedicated user group will be decided after the lab test in 2015. Based on the outcome of the test the decision will be made jointly with the production and programming department at RBB.

There will be a core test panel from 10-20 hard of hearing users. They will be interviewed at regular intervals through online questionnaires at least for several months to evaluate the sub-pilot service in detail.

Assuming that the sub-pilot will be an open service for all RBB viewers, there will be an open online questionnaire available through the RBB website for all interested users. The interested associations will also be involved and distribute a link to the questionnaire.

4.1.4. Validation Method

For this sub-pilot, RBB/IRT will organise an introductory meeting for all users from the dedicated test panel. During the pilot phase each test user will fill out online questionnaires, weekly or at least every second week. At the end of the pilot RBB and IRT will carry out closing interviews with all users from the test panel. This will be followed by a closing event for participants.

Assuming the Clean Audio sub-pilot will be openly available in Germany, we expect to gather a larger amount of quantitative data about the general usage and also some feedback from the interested users, especially through the good contacts to the associations.

4.1.5. General Set-Up

The test application will be realised as a browser based application.

The technical realization of the CA pilot will use a server-based approach (see section 6.1). This means that pre-produced CA content would be made available to the user (e.g. by allowing him to choose from different CA versions which are delivered by means of adaptive streaming using MPEG-DASH).

At RBB, there are the following distribution possibilities for providing Clean Audio:

a) Online media library online (PC)

b) HbbTV media library (Smart TV)

c) Online Live Streaming, possible only for pre-produced programmes (PC)

The current plan is to offer prepared audio-mixes in the media libraries. The parameters for the different audio mixes will be predefined based on the lab test in early 2015. The service and especially the chosen parameters will be re-evaluated with a core test panel of about 10 to 20 users.

The Clean Audio service could be integrated in above mentioned existing platforms or there could be an extra Clean Audio application. More specifically, the CA function could be integrated in the existing players, online and HbbTV, as known from the subtitle button. Next to the UT (subtitle) could be added a CA for Clean Audio.

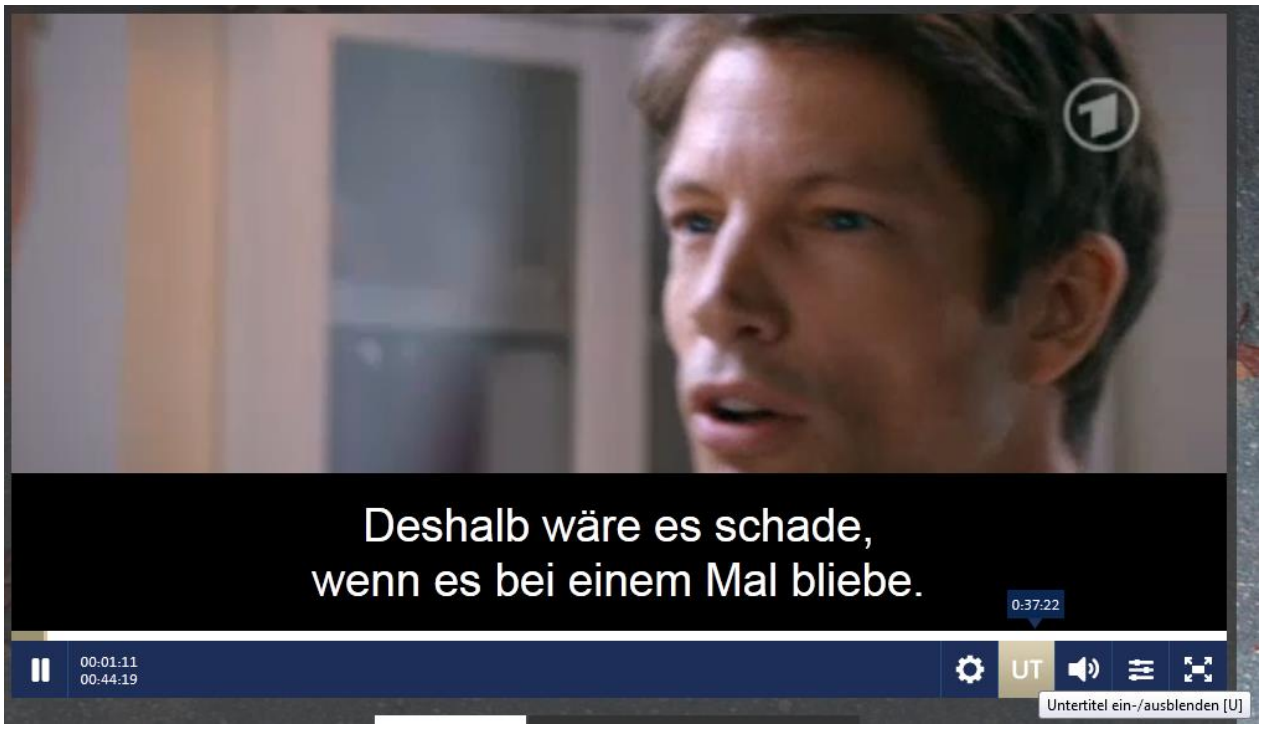

Figure 18. Screenshot online media library video player with subtitle functionality.

The other possibility would be to develop a special Clean Audio application for the target group, in which only Clean Audio versions would be available.

4.2. Clean Audio – Sub-Pilot in Spain (Catalonia)

4.2.1. General Set-Up

The foremost objective from TVC and UAB is to implement tests with users in lab conditions of Clean Audio content in Spain. IRT will assist in generating the CA content from audio material from TVC. Depending on the lab tests results and quality of the enhanced audio it may be scaled up for a pilot. It should be noted that for all intents and purposes, an enhanced dialogue can be treated as an extra audio channel available to users (specifically the hard of hearing, but also others may profit). So, once the CA audio is available it can be added to already existing audio choices for users, presumably using the same GUI. For the generation of CA signals, TVC will provide audio content to IRT, which in turn will provide the CA content to UAB to perform lab tests. Source audio material is expected to be in 5.1 and/or stereo format and will most likely not have separate audio channels for dialogue and FX. TVC will study the lab test results to evaluate the suitability of eventually developing a CA service, also taking into account the availability of the CA generation component by IRT and its licensing.

In case the lab tests prove satisfactory and the open pilot is deployed, content will be published using TVC's existing HbbTV on-demand video application, TV3 *A la carta*. MPEG-DASH on HbbTV will of course only be available to HbbTV 1.5/2.0 devices. At a later stage, content can also be reused for other platforms if feasible.

In case of deployment of an open pilot, the following scenario is planned:

a) Ideally, the IRT application to generate CA will be used by TVC to create test content (prerequisite: the application can be installed without too much effort at TVC). The content itself will be selected by TVC for suitability and relevancy. Automation of the workflow is purposely left out of scope so the content will be created manually using IRT's tool.
b) Depending on distribution details, MPEG-DASH feature support and HbbTV market deployment in Spain, the CA content will be made available at TVC's HbbTV on-demand application.
c) Availability will be seamless and transparent to users, that is, CA will be one of the audio options presented to viewers, who will select or not the CA option depending entirely on their needs and preferences.

Audience data will be gathered using in-app marking so results on CA content consumption can be collected, aggregated and analysed.

4.2.2. Status of work at M12

Suitable test content has been identified at TVC, ready to be processed as Clean Audio by IRT.

4.2.3. Timeframe

The timing largely depends on the availability of the CA implementation by IRT in two phases: user tests and possible open pilot. It is expected that IRT will be able to process test content by TVC by early 2015 (possibly needs to be matched with the lab test at RBB). UAB has a PhD student starting in April 2015 who could take care of user tests based on this test content.

4.2.4. Test Users

Lab tests are planned to be performed with 30 users. Users targeted in these CA tests will be "the aged" a population which is neither tagged as deaf nor blind, but are both. They are the highest population of people with accessibility needs in Europe. Users will be gathered from the Institute of Well Ageing at UAB.

4.2.5. Validation Method

Formal validation of the lab tests will be done by UAB using focus groups and the following tools:

- General questionnaire
- Interviews

In the case of the open pilot, TVC will use the existing production in-app marking facilities to ascertain the user consumption of the AC content. The TV3 *A la carta* service already gathers media consumption from the audience and the pilot will exploit the features already in place in this application.

4.3. Audio Description – Sub-Pilot in Spain (Catalonia)

4.3.1. General Set-Up

TVC already produces Audio Description (AD) using either spoken audio or in some cases Text-to-Speech, even though the service only deals with broadcast. For HBB4ALL the idea is to enable this technology in the online space, specifically in the case of HbbTV by leveraging MPEG-DASH, though results can be applied to other delivery and consumption platforms such as HTTP download, mobile, etc. Within the scope of the project, this objective is fulfilled in this manner:

a) Generation of AD using Text-to-Speech in Catalan from written text (usually subtitles)
b) Digital workflow changes needed to add this content to online publishing, including metadata handling, encoding, publishing and audience data gathering.
c) Playback of AD content using MPEG-DASH is also a requirement, other platforms being optional.
d) Feasibility of enhancement of the Text-to-Speech or regular AD using Clean Audio technologies, depending on tests within objective B-1
e) Application of enhancements in AD generation (Text-to-Speech or spoken) from recommendations issued by UAB given their lab tests and research.
f) Provision of test content to UAB should they need it for tests.

4.3.2. Status of work at M12

Requirement work is complete. Overall AD content workflow changes under analysis. AD content generation tests are underway. MPEG-DASH multiple audio support in development.

4.3.3. Timeframe

Analysis, design and development at TVC started September 2014, with the following planned milestones: AD asset generation by December 2014, followed by beta testing of AD publishing, metadata generation and content playback by January 2015.

4.3.4. Test Users

In the case of the open pilot Audio Description testing, potential test users are any users having an HbbTV 1.5 capable device within the region of Catalonia. Increasing the user base any further could only be achieved by implementing the fall-back solution of enabling progressive download of the same assets with embedded audio tracks and expanding the supported devices to HbbTV 1.1. This is undesirable due to the required replication of content so the initial target user base will be HbbTV 1.5 devices. In summary, in the case of AD an open pilot is planned and the whole of the Catalan region is the potential target (still taking into account HbbTV 1.5 market penetration), though effective potential users will be specifically those who benefit of AD features.

Additionally, tests will involve 30 Spanish and 30 UK blind users checking under lab conditions reception of ADs using Text to Speech technology.

4.3.5. Validation Method

User tests will be validated using existing audience activity gathering mechanisms already deployed in TVC's HbbTV service (in-app Javascript event gathering using Adobe Omniture).

4.4. Other languages and language learning – Sub-Pilot in Spain (Catalonia)

4.4.1. General Set-Up

Given that offering content in different languages is basically an instance of presenting different audio options to users, TVC will test and deploy in HbbTV the option to choose an audio track in a different language, wherever the content is available. Basically, this service will exploit all changes to fulfil objectives B-1 and B-2 done to support multiple audios in HbbTV and MPEG-DASH, exploiting the fact that TVC has a plethora of content in different languages. The service is expected to be part of the sub-pilot and finally put into production, publishing the content in multiple languages available to TVC. In this case, the experience can also be extended beyond HbbTV into other platforms if deemed cost-effective.

4.4.2. Status of work at M12

Requirement work is complete. Overall multiple language content support workflow changes in development. MPEG-DASH multiple audio support in development as in the case of section 4.3.2.

4.4.3. Timeframe

Start of the Analysis, design and development at TVC started in September 2014, with the following planned milestones: foreign language asset generation by December 2014, related metadata generation and publishing by December 2014, followed by publishing and playback of foreign language audio in HbbTV deployed live by January 2015 and beta testing of AD content playback also by January 2015.

4.4.4. Test Users

As in the case of section 4.3.4, intended tests users are the whole of the HbbTV user base in the region of Catalonia (taking into account market penetration and HbbTV 1.1 vs 1.5 installed bases). Unlike AD, foreign language is targeted at a broader audience and is expected to have more usage.

4.4.5. Validation Method

As in both previous objectives, validation of the pilot will be made using already existing audience activity measurement tools at TVC.

5. Roadmap

This chapter describes the Pilot-B roadmap at its current status. In section 5.1 an overview of the main timeline is given; this will be reviewed and adapted by the partners in the further course of the project. The further sections summarise the activities carried out and/or planned for each period; please refer to the other sections in this document for further details on activities and results from the first project period.

5.1.Main Timeline

Period	Activities
Dec 2013 – Aug 2014	Analysis of State of the Art and of Status Quo. Analysis of audio production workflows at TVC and RBB Initial definition of service pilots and sub-pilots Evaluation of available signal processing technologies for CA generator, technical testing, first draft implementation. Initial tests on playout system adaptations. Gathering of End User requirements Initial technical definition of audio services and potential workflows for operational phase.
Sep 2014 – Mar 2015	Improvements to CA generator; implementation of CA generator software component. Continue work on end user requirements. Initial prototyping of playout components. Finalise technical definition of audio services / workflows for operational phase.
Apr 2015 – Jun 2015	Finalise service pilots and sub-pilots definitions Finalise CA generator software component Integration of service components in workflows and playout systems Final preparations for Operational Phase
Jul 2015 – Dec 2015	Start of Operational Phase (T4.3) – final adjustments and improvements on service components
2016	Continuation of Operational Phase Evaluation of pilot and sub-pilot results (T4.4)

Table 3. Overview of Pilot-B timeline.

5.2.December 2013 – August 2014

The focus of this period was on obtaining requirements, covering preliminary tests and analysing the options for implementation of the envisioned audio services:

- Analysis of results of former projects and research activities; analysis of existing workflows and audio services offered (ALL)
- Detailed overview of service pilots and sub-pilots planned in each region (ALL)
- Collection of contacts for test panel (RBB)
- Selection and adaptation of audio signal processing technologies for a first (manually operated) version of the CA generator. Subsequent technical testing and processing of source material for user tests (IRT).
- With respect to requirements on CA implementation: preparation and carrying out of Focus Group meeting with accompanied CA user tests (RBB, IRT). Subsequent improvement of CA implementation and lab testing (IRT).
- Preliminary user tests (TTS, AD and language learning) for obtaining of requirements (UAB). Definition of stimuli to be used in user tests; definition of tests (UAB, TVC).
- Definition of the technical implementation of the audio services for the sub-pilots (IRT, RBB, TVC) in a draft version.
- Initial testing of MPEG-DASH multi audio encapsulation (TVC). Required as TVC currently does not have multi audio capabilities on HbbTV on production.
- Identification and evaluation of the shortcomings in terms of workflow and the modifications needed. (TVC)
- Collection of formal requirements, user stories and definition of mockups. (TVC)

5.3.September 2014 – March 2015

This period focusses on first (draft) implementations and narrowing down the implementation options for the operational phase. In parallel, further testing is performed.

- Implementation of draft version of CA generator software component (IRT). Preparation and carrying out of additional lab tests (RBB, IRT) and subsequent realisation of final component. Definition of approach to integrate Clean Audio (RBB, IRT).
- Definition of more precise target group with support of specialists (hard of hearing is a very big range). In addition contact with hearing aid acousticians. Planning of large in-house tests (RBB).
- Definition of suitable content for CA generation (RBB, TVC).
- Continue work on end user requirements and TTS / AD tests. Selection of stimuli to be used in user tests; carrying out the user tests. (UAB, TVC)
- Finalise technical implementation definition for audio services (e.g. target platform(s), playout modality etc.) in operational phase.(RBB, TVC, IRT)
- Initial work on implementation of HbbTV application / GUI (RBB)
- Testing of MPEG-DASH distribution of multi audio (IRT)
- Starting multi audio HbbTV distribution and AD transformation for multi audio on VoD HbbTV (TVC).
- Initial work on multiaudio MPEG-DASH distribution (TVC).
- Internal prototyping of HbbTV VoD (TVC).

5.4. April 2015 – June 2015

This period will focus on the final planning of the operational phase and the implementation of the services to be used in the pilots and sub-pilots. Most likely, further (user) testing will be carried out to support further improvements.

- Definition, final preparation and planning of service pilots and sub-pilots (ALL)
  - Recruitment of test panel
  - Questionnaires
  - Organisation of opening event with core test panel (RBB)
  - Planning of support (RBB)
- Finalisation of CA software component based on labtests feedback (IRT)
- Put workflows for multi audio services into place, integration of technology (RBB, TVC)
- Dissemination and communication with user associations (RBB)
- Briefing of editorial office for viewers (RBB)
- Multi audio HbbTV rollout, deployment of on-air pilot in Catalonia. Continuation of work on multi audio on live HbbTV. (TVC) User testing. (UAB)

5.5. July 2015 – Dec 2015

The operational phase of Pilot-B will start; the exact timeframe for running the sub-pilots is currently under discussion (T4.3).

- Roll out of service sub-pilot in Berlin-Brandenburg (RBB)
- Collection of data (RBB)
- Support especially of the core test panel of 10 – 20 users (RBB)
- Further testing / software updates (IRT)
- Preparation of evaluation (RBB)
- On-air rollout of CA on HbbTV if internal CA tests are successful and feedback is positive. (TVC)
- Begin of work on MPEG-DASH live distribution (TVC)
- Live support internal tests and on-air rollout if tests are successful.(TVC)
- Feed tests results to Quality Metrics deliverable (UAB)

5.6. 2016

Continuation of Operational Phase (T4.3) and evaluation and validation of the results of pilots and sub-pilots (T4.4). Processing of recommendations.

6. Service components to be adapted, integrated and tested

This chapter describes the developments of (technical) components specifically required to support the operational phase of Pilot-B and to establish the foreseen audio services.

6.1. Clean Audio generation

At the time of writing the project proposal, the HbbTV 2.0 standardisation progress was still quite open and it was not clear whether a client-based or server-based approach for the generation of a CA signal would be the preferred implementation for HBB4ALL. During the first 12 months of the project, the HbbTV 2.0 standard evolved and it is now clear, that the client-based approach cannot be realized with HbbTV 2.0 as essential features are missing for such an approach (see also section 3.3.1). Hence, it was decided to implement the CA generation based on a server-based approach, which means that pre-produced CA content will be transmitted to the user.

As the pilot aims at the automatic generation of the CA versions, a tool will be implemented which satisfies this requirement. Figure 19 and Figure 20 show the processing chain of the tool for both modes (basic and advanced). The basic mode will mainly enhance the relation of speech (in dB) to the rest of the audio signal.

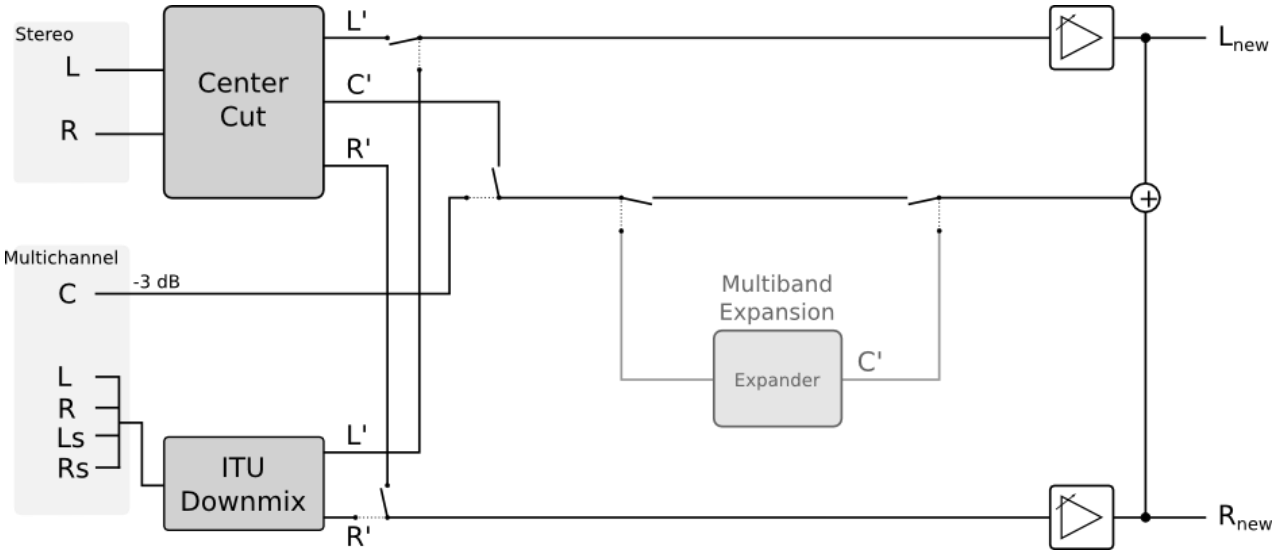

Figure 19. Processing chain of the Clean Audio generator (basic mode).

The first step is to separate the speech or dialogue from the rest. If only stereo content is available, the so called 'Center Cut' algorithm is applied to the input feed. This algorithm uses a correlation analysis to separate the middle of the stereo panorama (correlated parts of the audio signal in both channels). Since the speech within a typical stereo signal should be mixed to the middle of the stereo panorama, this is a quite useful and effective approach to separate speech from stereo content.

In a next step, the volume of the modified L' and R' channels is turned down by applying a negative gain factor (e.g. -9 dB). Finally, the newly generated Center signal C' is added to the reduced L_{new} and R_{new} channels.

For 5.1 content, there is no special speech extraction step necessary, since the Center signal (which is in most cases speech only) is already available. L' and R' are generated by applying an ITU downmix on the L, R, Left Surround (Ls) and Right Surround (Rs) channels of the 5.1 feed.

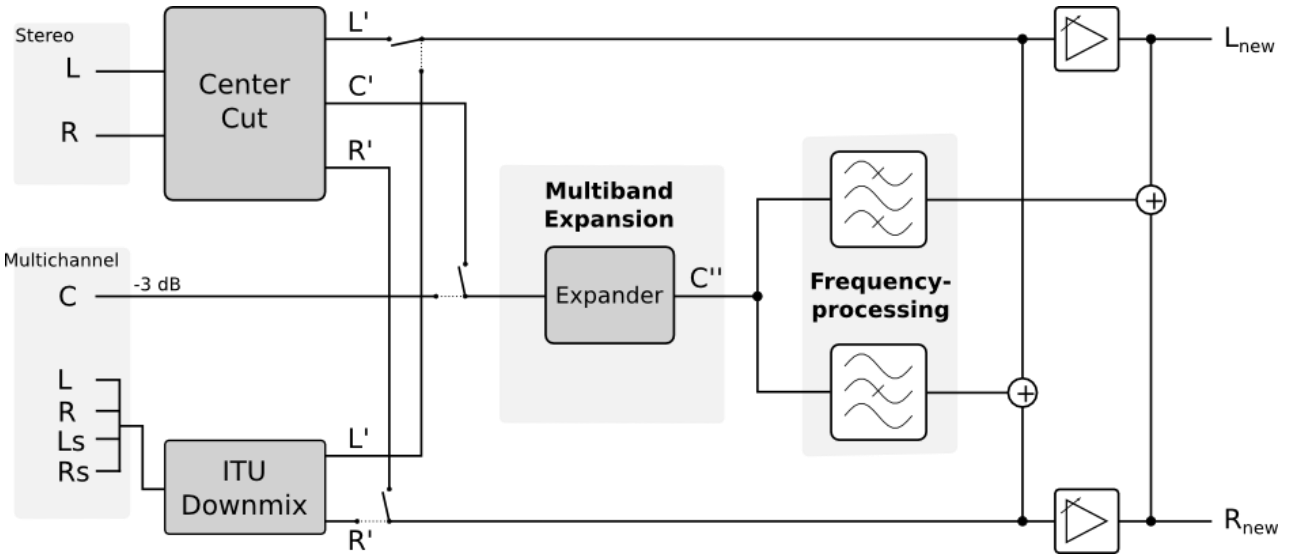

Figure 20. Processing chain of the Clean Audio generator (advanced mode).

The advanced mode of the CA generator follows the same approach for the speech extraction as the basic mode. But – in contrast to the basic mode – two additional processing steps are applied to the Center signal C (see Figure 20). First, if the dynamic (relation of slightest and loudest level) of the audio signal is too small due to heavy compression during the production, it will be enhanced by expanding the speech frequencies. In a subsequent step, the signal is filtered with a bandpass and bandstop filter. The bandstop (currently 20 Hz – 1.6 kHz and 8 kHz – 20 kHz) filtered part of the Center will then be added to L' and R' and afterwards the volume of L' and R' (including this inverse filtered part of the Center) will be reduced (as in the basic mode). The bandpass (currently 1.6 kHz – 8 kHz) will be added with original level to the reduced L' and R'. The purpose of this additional frequency processing is to emphasize the frequency range of the signal which is especially important for the intelligibility of speech.

6.2.Enhanced AD generation

Based on automatic Text-to-Speech generation AD content will be automatically generated (automatically generate AD with Text to Speech, Spoken Subtitles, combination of both). Any enhancements based on recommendations issued by UAB given the lab tests and research can be included in the existing system if required and feasible.

An audio mix of AD and the regular audio track is to be tested, due to workflow differences between broadcast AD and intended IPTV AD. In order to satisfy the HbbTV VoD workflow requirements the AD generation service will be improved, unifying in only one audio file including the whole AD in the timeline and including regular audio channel audio level changes for ease of listening.

6.3.Multiple audio asset generation

The digital media delivery workflow will be adapted so the multiple audio content is either a) generated separately but tagged as the same content or b) generated within the same digital 'asset', ensuring the publishing platform does not mistake what is actually one piece of content having multiple audios as separate discrete media entities. For details please refer to chapter 3.2.3 of D2.3.1 [2]. AD mixing generation analysis is ongoing to ensure the correct audio levels are generated.

6.4.HbbTV audio options selector

An HbbTV player with audio switching capabilities is going to be tested and included at the HbbTV application. For details please refer to chapter 3.2.4 of D2.3.1 [2]. User requirements are completed in this area and pending development once assets are ready and published.

6.5.Playback of audio content using MPEG-DASH

The CA and AD content is planned to be included as complementary audio channels at the HbbTV player. The components required for the generation of the MPEG-DASH assets, as well as for the selection of various audio streams in the end user device (MPEG-DASH "player") are being tested separately before being integrated. TVC and IRT bilaterally coordinate this activity to achieve a common approach and solution for using MPEG-DASH based audio within the HbbTV environment (specifically also in the HbbTV application).

Effective MPEG-DASH multiple audio support is still unclear at the time of writing, and subject to combined efforts by IRT and TVC. The option of generating multiple assets is still open as a fall-back solution.

6.6.HbbTV device and version detection

As for Pilot A, a JavaScript component is going to be developed to detect the device capabilities, specifically its HbbTV standard version to ensure any features used are compatible with the viewing device. For details please refer to chapter 2.2.8 of D2.3.1 [2]. Implementation is proving relatively straightforward and can be integrated within the existing TVC HbbTV application. If required, the RBB and TVC approaches can be compared and shared if necessary.

7. User tests

To support the requirements finding process and the fine-tuning of services and service components, so-called "preliminary" user tests are and will be carried out. Some have and will be done before the actual start of the service pilots; additional user tests may be carried out also during the operational phase. Within HBB4ALL user tests have been defined as tests of "partial services", e.g. testing a limited set of the envisaged service features or parameters. Some may be technical or lab tests – focussing on the correct implementation or functionality –, some are set up with potential target users to get their feedback as input to the service requirements.

For Pilot-B already within the first project period a variety of preliminary user tests have been performed. A description of these and their intermediate results are presented in section 7.1. An overview of additional preliminary user tests currently planned for the second period is given in section 7.2.

7.1. Preliminary user tests (1st period)

7.1.1. Clean Audio

7.1.1.1. Focus Group, Berlin

As agreed at the kick-off meeting, IRT and RBB held one focus group meeting in July 2014 (as the 1st step in Task T4.1) with 12 users from the target group (acquisition via associations for the disabled in Berlin-Brandenburg) to define basic requirements.

The event lasted for approximately two hours at RBB in Potsdam. After the welcoming procedure the users were divided in two equal groups. One group began with the audio sample test, the other group with the questionnaire. Then the groups switched and the first group completed the questionnaire while the second group did the audio sample test. Afterwards the two groups came back together and there was an open and fruitful discussion.

All subjects of this test have a very severe hearing impairment and are wearing hearing aids. The analysis of the questionnaire showed that all participants have problems regarding the speech intelligibility for almost every type of TV content. As most problematic factors of audio signals were identified:

- Background music
- Noise
- Effects
- Grubby articulation
- Dialect

Another, often mentioned request was the possibility to adjust the level of speech and background noise (music, atmo, effects) to the personal demand.

Figure 21. Focus Group at RBB.

At a first glance it can be assumed that the planned approach and structure of the Clean Audio generator targets the identified problems or requirements. The first three factors from the above list can be addressed, whereas it is impossible to implement any improvements for the last two factors (articulation and dialect). As the questionnaire is also available as an online survey for a wider audience, the detailed analysis of the questionnaire results will be delivered along with the results of the online survey.

The listening test was conducted in real world environment: IRT and RBB made sure that the acoustic of the room was adequate ('living room quality') since the test stimuli were played back via loudspeakers. Moreover, several subjects conducted the listening test at the same time. Figure 21 depicts the set-up at RBB premises. The stimuli were carefully chosen to test different input formats (stereo and 5.1) and both modes of the generator (basic and advanced) along with different content situation. For the latter, one can distinguish between the contents by the availability of clean speech in the Center channel of the 5.1 signal. For two contents, the Center channel had speech only ('Center clean') and for the other two, the Center channel also contained some effect, atmosphere and music ('Center not clean'). Moreover, two different mixing levels were also tested (-9 dB and -15 dB). These properties resulted in 24 combinations which had to be evaluated.

After an adequate familiarization with the test procedure and the content, the listening test started and lasted approximately 30 minutes. After 15 minutes, the test was interrupted for a short break to minimize the stress for the test persons.

During the listening test, the test persons could switch between a reference signal and a modified signal. The test persons were instructed to rate both the speech intelligibility as well as the listening experience of the modified version compared to the reference signal. The order of the stimuli was chosen randomly.

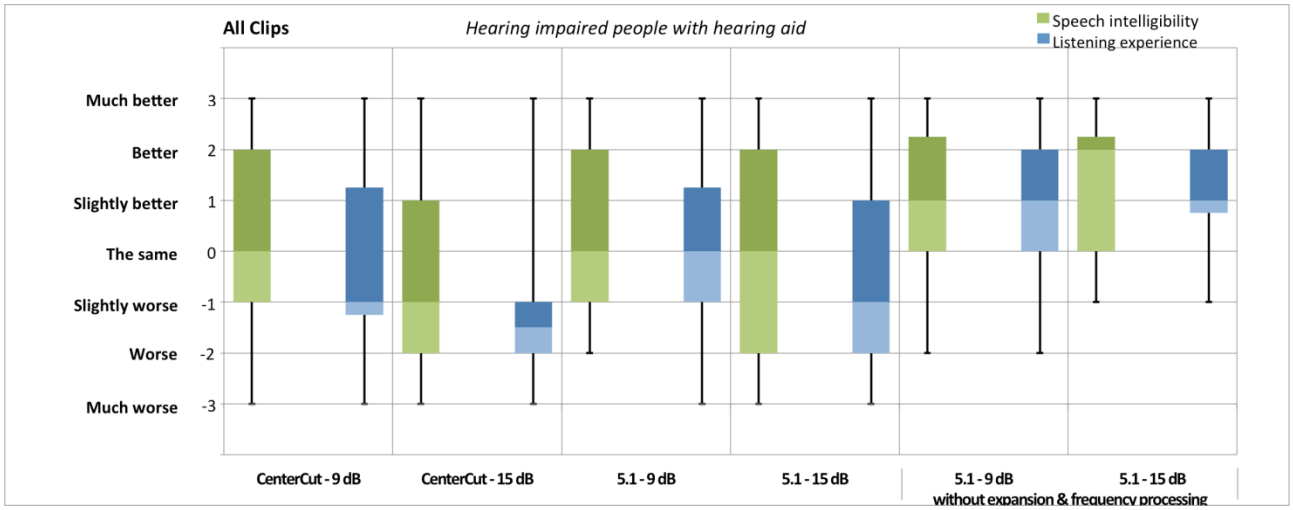

Figure 22. Results for all stimuli for the test group at RBB premises (hearing impaired people with hearing aid).

Figure 22 shows the overall results of the listening test for all stimuli ("clips"). The values are illustrated as box plots. The black line depicts the range (minimum and maximum rating) of all ratings. The overlap of dark and light green shows the median value. The upper, dark green end of the box illustrates the 25 % percentile and the lower end (light green) the 75 % percentile of all ratings.

The stimuli called 'Center Cut' are versions generated with the advanced mode for a stereo input feed. '5.1 - 9dB' and '5.1 -15dB' means that the 5.1 input feed was processed in advanced mode. The two stimuli on the right side of the Figure 22 were generated with the basic mode for a 5.1 input feed.

It can be seen that the most efficient processing could be obtained for the basic mode stimuli (without frequency and dynamic processing). The test persons rated the speech intelligibility over-all with 'slightly better' or 'better'. The listening experience also was rated positively which leads to the conclusion that for the target group with hearing aids the basic mode should be preferred.

7.1.1.2. Labtest, Munich

The test was repeated with other hearing impaired people at IRT premises in July 2014. The test persons in this group have a slight to middle severe hearing impairment and do not wear a hearing aid. The test persons indicated that they have problems regarding the speech intelligibility of TV signals. The number of test persons was ten. To get an impression of the severity of the individual hearing impairment of each person, an audiogram was generated beforehand with sufficient timely distance to the listening test. After the post-screening and the analysis of the audiogram, three persons had to be excluded as their impairment affects only one ear. Hence, it can or must be assumed that the sane ear compensates the hearing impairment.

The test procedure was identical with the one performed at RBB premises. Figure 23 shows the set-up at IRT premises. Since the number of test persons cannot guarantee a full statistical validity, the results of this

listening test should be seen as an orientation for further developments and for the second lab tests planned for February 2015.

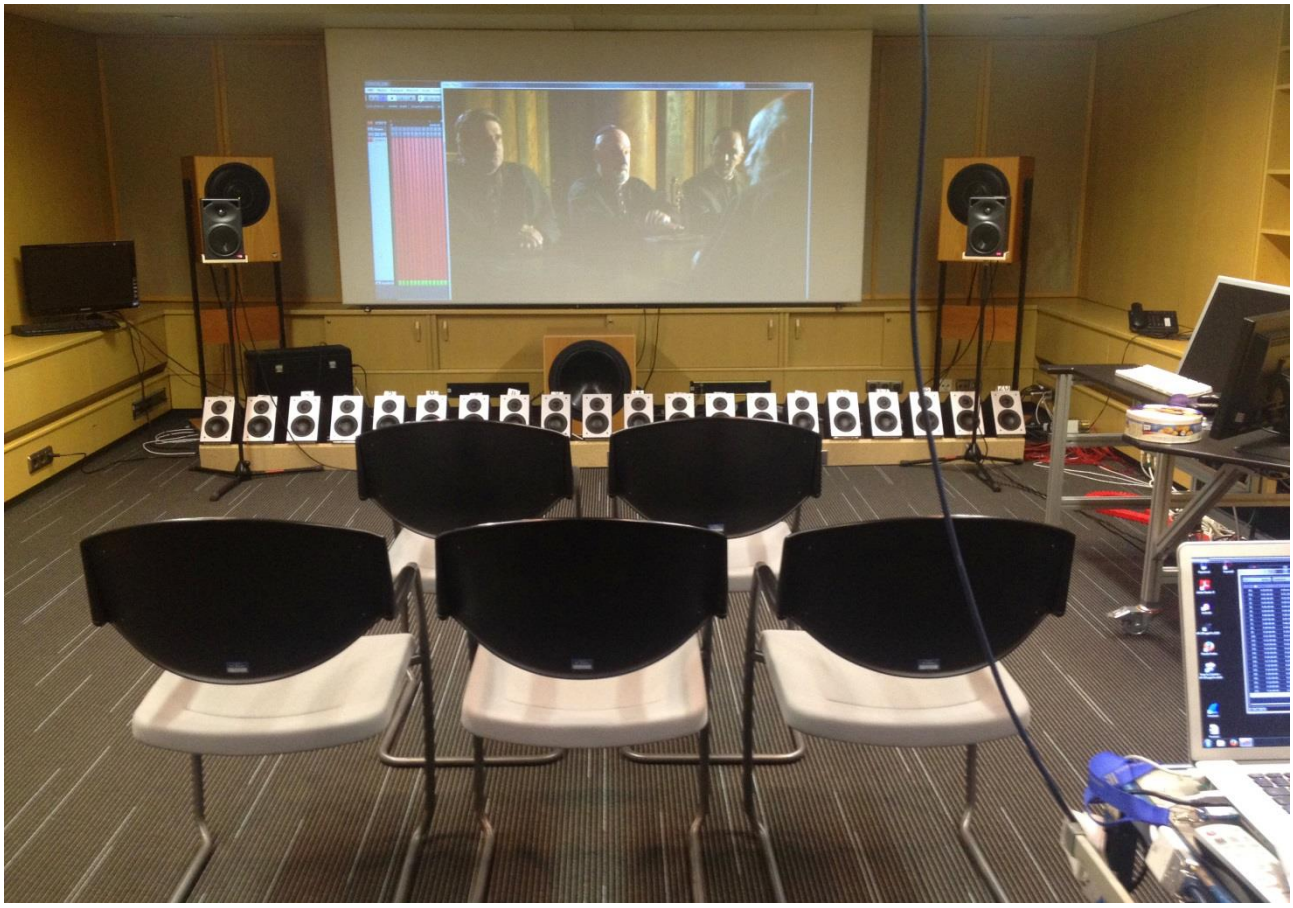

Figure 23. Listening test set-up at IRT premises.

The test results were not as clear as from the test conducted for the hearing impaired people wearing a hearing aid. It was observed that the property 'Center clean' or 'Center not clean' influenced the ratings of this group more than the group with hearing aid. Therefore, the results were analysed taking this distinction into account.

Figure 24 illustrates the results for the clean Center signal stimuli. The best rating was reached for the stimuli from 5.1 input format with mixing level -9 dB including frequency and dynamic processing (advanced mode). The average result was rated with 'better', though the listening experience suffered slightly for this type of processing. Good results could be also reached for the stimuli without frequency and dynamic processing (basic mode). The listening experience of the basic mode stimuli also suffered slightly from the signal processing. Nevertheless it is assumed that the better speech intelligibility outweighs the slightly worse listening experience. In interviews after the listening test the test subjects confirmed this assumption.

Figure 25 depicts the ratings for a Center signal with noise. For no single version of the stimuli the result could be identified as the best. The listening experience seems to be preferred for the basic mode versions. Since one of the two contents had very much noise in the Center signal, it can be seen as an extreme example

which will probably rarely occur in productions. The results for a second content with less noise in the Center again showed an improvement regarding the speech intelligibility.

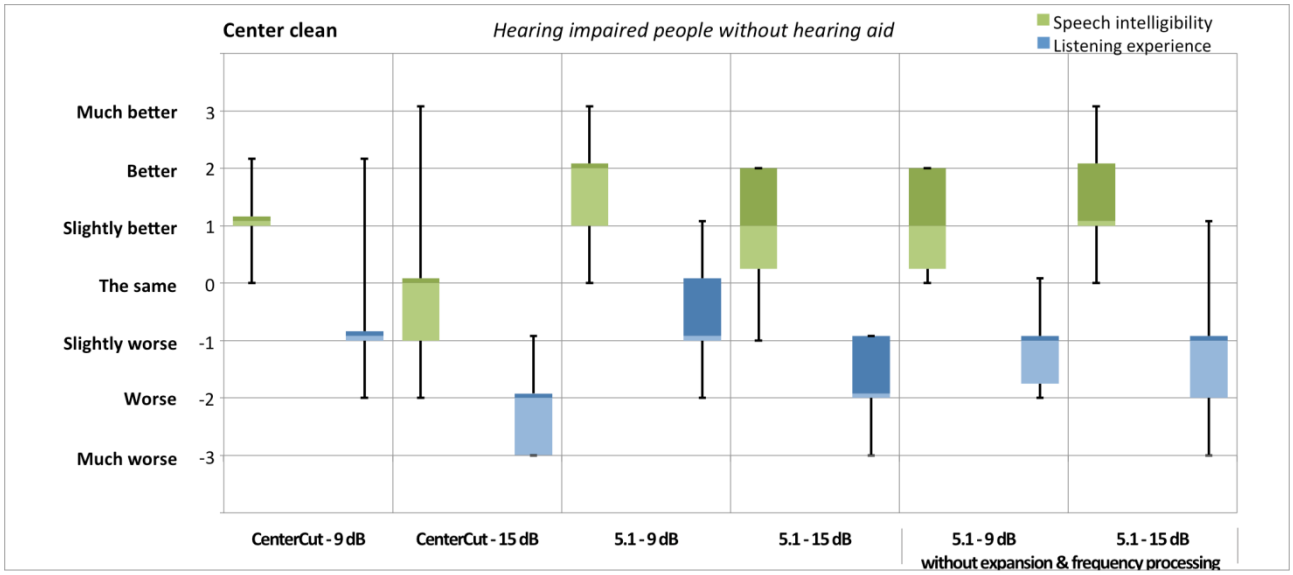

Figure 24. Ratings for hearing impaired people without hearing aid for content with clean Center channel.

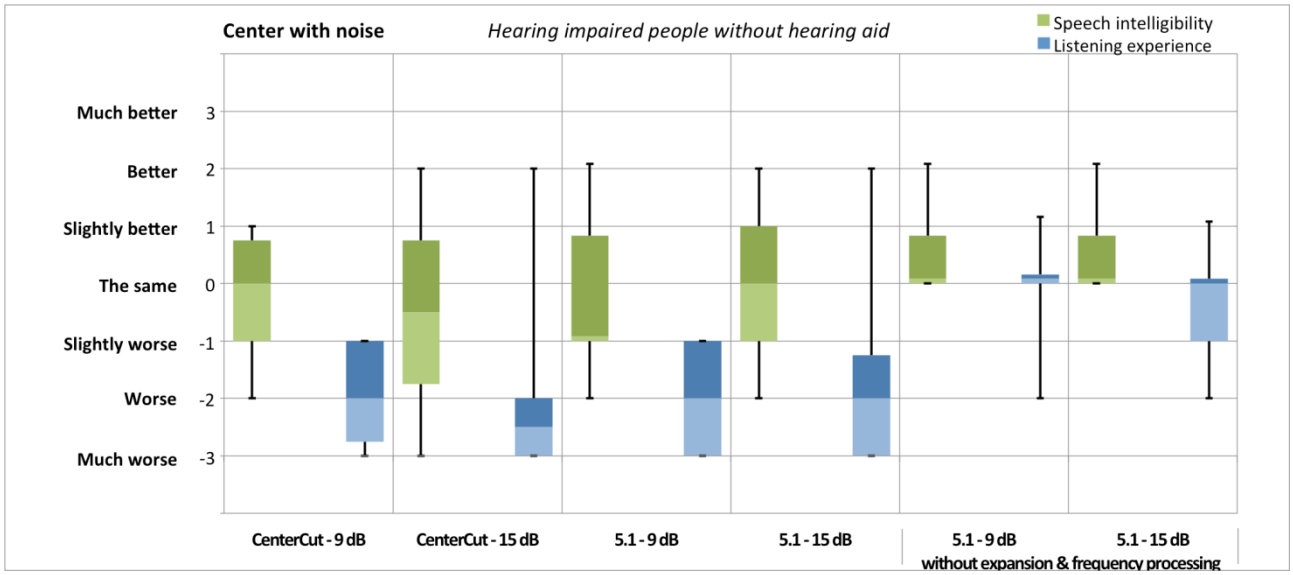

Figure 25. Ratings for hearing impaired people without hearing aid for content with noise in the Center channel.

7.1.1.3. Conclusions from German Focus Group and Labtest

Summarising, the intermediate conclusion is that 5.1 content is the preferred input format as the results for either 5.1 with basic mode or 5.1 with advanced mode had the highest ratings in our tests. If 5.1 content is available, the Center speech signal should be clean as this leads to the best results. It can be also derived that the signal processing for the pilot should potentially distinguish between people wearing a hearing aid and people who are not. In interviews and discussions with the test persons with hearing aid, it became clear that these people mostly benefit from the basic mode processing. Moreover, experts from a hearing aid manufacturer as well as one specialist for the testing and setting of hearing aids confirmed that the most efficient improvement can be reached by enlarging the speech to background noise ratio. Additional signal processing as it is applied with the advanced mode can even have negative effects on the intelligibility as the hearing aids already have a similar processing included. This would lead to an undesirable double processing what can be also seen in the ratings for the advanced mode of the test at RBB.

Since the pilot will be implemented as server-based approach, a continuous adjustment of the mixing level between speech and background cannot be realized without greater effort and technical problems or even limitations. Hence, we plan to offer several pre-produced versions with different mixing levels. This is a compromise between full personalization and technical feasibility and still supports the project goals. The exact number and gain factors of the versions have to be defined and tested in the second lab test in February 2015.

7.1.2. Audio Description

7.1.2.1. Audio description and Second Screen Application

As part of the UAB participation in the EU project HBB4ALL, UAB along with TV3 set up some tests to gather end user reception on usability and accessibility for the app ArtAccéss, developed by the company S•Dos for the Catalan Government[7]. During the Sitges Film Festival 2014, ArtAccéss for AD was tested.

The app had following functionality:

- Download, store, and play adapted content (subtitles and/or Audio Description).
- Synchronize at any time during play, through sound stamp.

[7] http://www.sdos.es/accesibilidad/

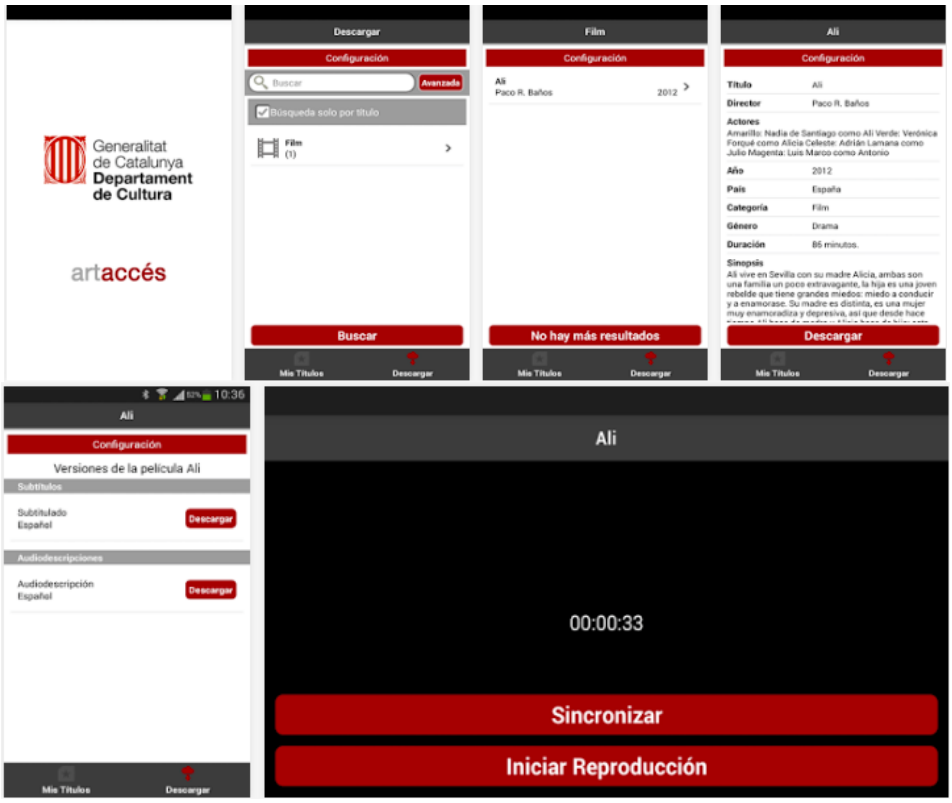

Figure 26. Sequential screens for ArtAccéss Application.

The Catalan Government made possible the tests taking care of the content services, contacting user associations (ACCAPS, FESOCA, ONCE and ACIC), paying for tickets to enter, and also booking a special room for interviews. Public Catalan TV3 and UAB prepared tests and questionnaires.15 users were present for the AD tests (7 men and 8 women). Out of them 10 were blind and 5 partially sighted. There was also a control group consisting of 5 sighted persons.

Figure 27. Study Participants.

Results

In general, the blind and partially sighted users evaluated the application positively. They especially liked the fact that they could download the AD file at home. They underlined that it was comfortable for them to use the application. The cinema was full of people and they did not receive any complaints as far as the use of the application was concerned. The stability of the application and the fact that it did not fail throughout the film were listed as its strong points, whereas interferences with other applications were deemed its weak point. Also, a definite advantage of the application is the fact that its display screen is black, which does not consume much battery and does not bother other cinema goers. Bearing in mind all this, it is considered that the application has a big potential. As soon as improvements have been introduced, it could be tested in a more representative context, e.g. with other films, other language versions and perhaps with other devices like tablets or different models of mobile phones. More tests should also be carried out taking into account different types of user profiles.

Requirements on AD over secondary screens deducted from user tests

- Language of AD and language of audio-visual content must be the same

  AD should be in the same language as the content. In bilingual countries, such as Catalonia, special attention should be paid to keep the same languages for both content and its AD.

- Navigation screen for app should be carefully designed

  The user interaction with the secondary screen functions should be carefully designed taking into consideration usability principles, and always bearing in mind the user as "VIP" who uses a screen reader to navigate through the screens.

- Sound volume must be adjusted, since at times it is not high enough

  Sometimes the volume of the content is very high, for example at the end of a competition, when the winner is announced. In these cases the volume of the AD should be able to be higher.

- AD sound must be mono and not stereo, since AD is heard by only one earphone

  Users consume the content with only one earpiece where they listen to the AD, with the other ear they listen to the original content sound. AD should be mono and not stereo, since only one earpiece is used.

The above mentioned requirements were delivered to the application developers / manufacturers and relevant improvements are to be introduced. Further testing is to be scheduled with blind and partially sighted participants.

7.1.2.2. Quality of genre-specific AD

User tests were carried out to check the preferences of AD users for neutral vs alternative version of AD (see Background Information below). There were 15 blind and partially-sighted participants involved. Some tests have already been conducted in Poland (they took place in August/September 2014; results to be delivered by 31st December 2014).

In general, as intermediate summary, it can be said that the hypothesis has been confirmed: AD users showed a preference for the alternative version of AD (compared to a neutral AD version). The above seems to be quite an important finding as far as training of audio describers is concerned. However, in order for it to be incorporated in the official AD guidelines, further testing seems indispensable. If confirmed, the preference by users of non-objective AD is a ground-breaking result, which may alter AD content production.

Background information

For the purpose of the study, two AD scripts were prepared: one with emotive references taken from the screenplay and the other more detached, which follows the existing tradition of "what you see is what you say" rule. We verified whether a powerful and naturalist drama would render itself to this type of description and whether this type of description increased the target group's engagement in the film. In this context, we also touched upon the question of AD quality which is inextricably linked to users' experience and proposed a tool for measuring it, i.e. the ITC Sense of Presence Inventory (ITC-SOPI, Lessiter et al. 2001) [3] short form. Such questionnaires were handed out to participants after the screening and they included the evaluation of their emotional engagement and immersion in the content presented.

7.1.2.3. Sound mix in Audio Description

Given the different technical processes and requirements of sound and sound mixing depending on the channel through which it reaches the listener, these tests by UAB were restricted to sound quality in Audio Descriptions of pre-recorded content to be TV broadcasted. This restriction was due to two main reasons: 1) TV is the most common channel of audio-visual content consumption since 97% of households had a TV and regard it as basic hardware, 2) because the content needs to be delivered with sound levels, mixing, and adjusting, and it requires pre-recorded content to work with.

The hypothesis of these tests was that a "safe area" in terms of intelligibility can be defined to guarantee that the Audio Description narration is always heard correctly in any audio-visual product/service. An adequate sound post production process can guarantee that such a "safe area" is always maintained, to the same extent as an inadequate sound mix could seriously threaten the quality of any Audio Description.

Preliminary Conclusions

A manual adjustment of the AD narration volume levels during the postproduction process would imply an increase in the audio quality of the final result. This process, opposite to what proposed in the ITC guidelines, would keep the original soundtrack and effects intact, preserving thus the original features and atmosphere of the work. Having the AD narration volume pre-adjusted to the loudness level of each of the scenes where it will be inserted, would avoid abrupt atmosphere disruptions of quiet scenes and possible difficulties of understanding of the narration in noisy scenes.

In the future, this postproduction process would contribute to maintaining a consistency of the volume of the AD narration in relation to the original soundtrack all through the program/film described.

Results from these experiments will also be used in the composition of Quality Benchmarking (HBB4ALL deliverable D2.6.x).

7.1.3. Other languages and language learning

7.1.3.1. Audio Description as a tool for language learning

This study was carried out in Poland in the in the first half of 2014, with 36 blind and partially sighted participants and sighted students. The study aimed at examining the influence of audio described films centered around introducing new vocabulary on foreign language learning of school children with and without vision impairments. The audio-visual material employed for the purposes of the study included four episodes from educational animation series "Say it with Noddy in English", namely: "Hammer", "Moon", "Present" and "Paint". All of them were dubbed into Polish.

The key objective of the study was to examine whether educational movies with Audio Description (AD) facilitate foreign language acquisition and whether AD itself can help both blind and sighted children to better understand the content of such movies. The study also tested whether there are any differences in the results between visually impaired children and their sighted peers. Finally, the study looked into whether audio describing this kind of films could prove helpful for young viewers with and without vision impairments, and more importantly, whether such audio-visual materials could become an additional educational tool.

The study showed that the above assumptions were correct and it is worth introducing films with AD into the school educational programmes. The educational value of AD was confirmed, not to mention the integrating function of films with AD. For the achievement of objective B-3 these are important intermediate results.

7.1.3.2. Audio Description as a didactic tool

This study was carried out in Poland in the in the first half of 2014, with 55 blind and partially sighted participants. It involved classes of history, biology and physics with Audio Description (AD).

The aim of the study was to verify the possibility of using AD as an additional didactic tool in the education of the blind and partially sighted children. More specifically, the study meant to investigate whether and if so, to what extent adding AD enhance the comprehension of educational films among children and what could be the possible value of AD in educational settings. In addition, the underlying objective of the study was to raise awareness among teachers about the importance of AD as a means to achieve audio-visual accessibility. The filmic material employed for the purposes of the study was three episodes from the educational animation series "Once Upon a Time...", namely "Once Upon a Time... Man" (episode "America"), "Once Upon a Time… Life" (episode "Blood") and "Once Upon a Time… Discoverers" (episode "Newton").

After each screening, the children were invited to take part in a quiz competition, with the aim of testing their level of comprehension of the shown film. To gather children's responses the researchers employed innovative devices called clickers and supporting software from InterWrite PRS. Both the screenings and the questionnaires with clickers were greeted with much enthusiasm by children and they arouse interest and curiosity also among teachers.

The study revealed that AD in educational films helps to develop and retain specialised vocabulary. Moreover, it helps the children to follow the film, thus improving its overall comprehension. It is the researchers' assumption that AD, through its use of varied word choice, synonyms, metaphors and similes, can boost literacy of the blind and partially sighted children. Finally, a conclusion can be drawn that films

with AD could not only complement lessons, but also make them more enjoyable. For the achievement of objective B-3 these are important intermediate results.

7.2.Planned preliminary user tests (2nd period)

This section overviews the currently planned preliminary user tests for each of the objectives in Pilot-B.

7.2.1. Clean Audio

7.2.1.1. Expert testing (Spain)

In the case of Clean Audio, in addition to the tests performed by UAB, TVC plans to have informal qualitative testing of any results to be performed by TVC technical experts and audio production professionals. The results of these tests would, as far as appropriate, be fed back to IRT to further improve the quality of the Clean Audio generation.

7.2.1.2. Labtests (Germany)

Based on the defined basic requirements from the focus group in July we plan to perform an additional lab test with about 20-30 hard-of-hearing people in February 2015 at RBB. The objective of this test will be to define the mixing levels of the CA versions which will be generated and offered to the end users. These parameters will be the basis for the service pilot.

Each user will do a session of about 1,5 hours of testing at RBB. Following methods will be used:

- General questionnaire
- Listening test (comparable set-up as for the tests in July 2014)
- Interview

7.2.1.3. Labtests (Spain)

Clean Audio refers to audio providing improved intelligibility. It is targeted for viewers with hearing impairments, but can as well serve as improvement for listening in noisy environments like airplanes. Audio mix with fixed balance between dialogue and background is always a compromise and therefore finding the right mix can pose a challenge. Bearing in mind the above, end user tests will be carried out. They will focus on transmitting several separate audio sources and eliciting users response to them. Additionally, speech intelligibility listening tests will be performed.

Based on the defined objectives and from previous data results TVC and UAB plan to embark in new tests which will aim at examining how Clean Audio enhances foreign language learning, and learning in general. At present TVC and UAB are defining stimuli, tests and schedule.

7.2.2. Audio Description

7.2.2.1. Quality of AD with cultural allusions

Tests are planned to check whether users benefit from AD with cultural allusion and references. This will be checked through checking levels of engagement and comprehension in AD. The tests will be conducted in Spain, Italy and the UK, and include around 45 blind and partially-sighted participants (~15 participants per country).

Background information

Cultural allusions and intertextual features are recognized as important issues in Translation Studies, however little is known about their impact on blind and partially sighted AD users. This study aims to compare English and non-English ADs for the same film, with English and non-English audiences to see how much or how little these elements are picked up, and the subsequent impact on engagement and comprehension. It is expected that access to cultural allusion and intertextuality will increase engagement and comprehension such that the English audience reports greater levels of both than the non-English audiences.

7.2.2.2. AD quality: getting the priorities right

AD technical quality standards need being defined in order to guarantee a quality accessibility service. Since technical standards regarding sound requirements are not established in any of the guidelines issued up to now, a literature review on the current state of this issue will be made. If no clear standard can be found on the available documents on AD guidelines and recommendations, we suggest setting up an experiment to find out the receiver's preferences between two models of adjusting sound levels of AD narration:

- Lowering the background Original Sound Track (OST) so AD is always heard clearly
- Manually adjusting the AD so its sound level goes up and down in harmony with the OST

7.2.2.3. Audiointroductions

Tests are planned to check the preferences of AD users for human voice versus synthetic TTS delivery of the audiointroductions (AIs) and to elicit participants' interest in accessing AIs (see background information below). There will be a total number of 45 blind and partially-sighted participants involved (15 participants per country). The tests will be conducted in Poland, Spain and the UK in the first half of 2015.

Background information

Audiointroductions (AIs) – also known as introductory notes, show notes or programme notes – have been used in opera and theatre since the early days of AD. They are pieces of continuous prose, spoken by a single voice or a combination of voices lasting between 5 and 15 minutes. AIs aim to create a framework by which to understand the action; they have an information function providing relevant details such as running time, cast and production credits, as well as detailed descriptions of the locations, costumes and characters, and can convey a sense of visual style including camerawork and editing. Currently AIs are not provided for film or TV but a study by Romero-Fresco and Fryer [4] suggested blind people would welcome the option to download them as an audio or text file from a website. Alternatively, HbbTV can offer a good opportunity to

deliver audio introductions (AIs), as an alternative and complementary source of information which can be activated by the user at any time during the TV programme.

7.2.2.4. Impact of Clean Audio on Audio Description

Additional lab tests may also include the verification of the impact of applying Clean Audio in Audio Description. Using IRT's implementation of CA, TVC would apply CA to the AD channel so UAB could check in lab conditions the impact of applying such techniques in the Audio Description domain. This is seen as a 'target of opportunity', as both CA and AD are tested separately within the Pilot-B, the opportunity is there for UAB to perform experimental lab tests to combine both technologies. It is not known at present if this work will yield any significant results and the test at the moment is tentative.

7.2.3. Other languages and language learning

7.2.3.1. Impact of Clean Audio in language acquisition

Tests will be taken on board to check if Clean Audio improves language learning. These tests will be carried out with TVC. See section 7.2.1.3.

8. Results

This chapter summarises the progress of each objective in the active tasks (T4.1 and T4.2) of this period and separately on the component implementation.

8.1. Task 4.1 Progress

Objective	Progress
Evaluate requirements from broadcaster and end-user perspective	70%
Analyse Status Quo of the aimed audio services	100%
Define features for alternative audio services in operational phase	40%
Review and implement appropriate broadcaster workflows for the aimed audio services	50%
Consider various options for generation and distribution of Clean Audio	70%
Consider metadata modifications required for multi-audio asset production	60%
Realise automatic Text-to-Speech generation in Catalan language to automatically generate AD content	40%
Define success criteria for later user validation	20%

Table 4. Summary of Task 4.1 progress.

8.2. Task 4.2 Progress

Objective	Progress
Implement and integrate all technology needed based on the outcome of T4.1	30%
Conduct expert or friendly user viewings to gain feedback with respect to technical feasibility, perceived user value and business sustainability.	20%
Integration work to publish A/V assets and metadata	0%

Table 5. Summary of Task 4.2 progress.

8.3. Component Progress

Component	Progress
Clean Audio generator	50%
Enhanced Audio Description generation	20%
Multiple audio asset generation	30%
HbbTV audio options selector	20%
Playback of AD content using MPEG-DASH	20%
HbbTV device and version detection	80%

Table 6. Summary of WP4 component progress.

9. Conclusions

Work Package 4/Pilot-B has advanced well during the first project period. This document presents in detail the degree of accomplishment of the objectives for the initial two tasks of the pilot (T4.1 and T4.2).

- A large amount of preliminary user testing has already been carried out as well as technical feature testing and requirements gathering.
    - With respect to Clean Audio, the test results have been used to enhance the implementation of the Clean Audio generator and tune it for use in the pilots.
    - With respect to Audio Description, we have results confirming:
        - AD enhances learning in general;
        - AD enhances learning foreign languages;
        - AD on secondary platforms has a high degree of user acceptance.
- The technical service planning (target platforms, playout modality) has started; final decisions depend on further preliminary user tests.

Obtaining appropriate audio source material for Pilot-B has been doubly challenging due to the fact that 1) all rights have to be cleared for the content (such as in the case of foreign language audio) and 2) quality audio tracks for Audio Description and Clean Audio generation have to be available.

With respect to the original work plan, there will be some small deviations, specifically because of unforeseen delay in the publication of the HbbTV 2.0 specification (so that most likely no HbbTV 2.0 devices will be available for use in the pilot) and the details of the specifications itself, which were unknown at the time of writing the Description of Work:

- Live use case for pilot: very unlikely as the key “Multi-stream synchronisation” feature from HbbTV 2.0 would be required to implement a live use case where a TV program (sent via DVB) is enhanced by synchronisation of additional audio stream(s) delivered via IP (broadband). Therefore Pilot-B will focus on on-demand services.

HBB4ALL partners are looking into the possibility of realising this as a showcase based on early HbbTV2.0 prototype receivers.

- Clean Audio: early decision to focus on a server-side solution for the Clean Audio generator. For a client-side solution multiple (at least two) audio decoders would be required in the receiver, to allow audio mixing. However, most likely only a single audio decoder will be mandatory in HbbTV 2.0 so a server-side implementation is the only feasible option.

The main objectives for Pilot-B can still be achieved, focussing on HbbTV 1.0/1.1/1.5 and no further specific measures, actions or contingency plans are considered necessary at this stage.

10. References

[1] Maszerowska, A., Matamala, A. & Orero, P. (2014). *Audio Description*. New Perspectives Illustrated. Amsterdam: John Benjamins

[2] D2.3.1 – *Common Technical Components (I)*, HBB4ALL deliverable, July 2014

[3] Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). *A cross-media presence questionnaire: The ITC Sense of Presence Inventory*. Presence: Teleoperators, and Virtual Environments, 10 (3), 282-297.

[4] Romero-Fresco. P. & Fryer, L. (2013). *Could Audio Described Films benefit from Audio Introductions? An Audience Response Study*. Journal of Visual Impairment and Blindness 107 (4), 287- 285.

[5] Eberhard, M., Eberhardt, M., Ebert, E., Hergenröder, Prof. Dr. E. Kiefer, M., Matejka, D., Siegfried, A. (2014) *Sprachverständlichkeit im Fernsehen – Empfehlungen für Programm und Technik*. ARD/ZDF. Downloadable from www.irt.de/webarchiv/showdoc.php?z=NzE0MSMxMDA2MDE4I3BkZg