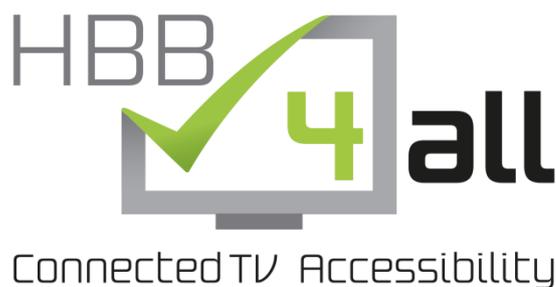


## D4.2

<b>Version</b>	1.00
<b>Date</b>	2015/09/30
<b>Author</b>	IRT
<b>Dissemination status</b>	PU
<b>Document reference</b>	D4.2



## D4.2 – Pilot-B Solution Integration and Trials

<b>Grant Agreement n°:</b>	621014
<b>Project acronym:</b>	HBB4ALL
<b>Project title:</b>	Hybrid Broadcast Broadband for All
<b>Funding Scheme:</b>	CIP-ICT-PSP
<b>Annex I reference version:</b>	2013/10/22
<b>Project Duration:</b>	2013/12/01 – 2016/11/01 (36 months)
<b>Coordinator:</b>	Universitat Autònoma de Barcelona (UAB)
<b>Beneficiaries:</b>	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 2017-2013' of the European Union.





CIP-IST-PSP-621014

[www.hbb4all.eu](http://www.hbb4all.eu)

D4.2 v1.0

Project no. 621014

**HBB4ALL**

Hybrid Broadcast Broadband for All

CIP- Pilot actions

Competitiveness and innovation framework programme 2007-2013

## D4.2 – Pilot-B Solution Integration and Trials

Due date of deliverable: 2015/07/31

Actual submission date: 2015/09/30

Start date of project: 2013/12/01

Duration: 36 months

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

Revision 1.00

Project funded by the European Commission within the Seventh Framework Programme		
Dissemination Level		
<b>PU</b>	Public	<b>X</b>
<b>PP</b>	Restricted to other programme participants (including the Commission Services)	
<b>RE</b>	Restricted to a group specified by the consortium (including the Commission Services)	
<b>CO</b>	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	2015/05/15	-	Draft document structure	IRT
0.11	2015/05/22	-	Overall structure adapted to Pilot-B	IRT
0.12	2015/05/26	-	Final structure of D4.2	IRT
0.2	2015/06/26	all	Initial input	IRT, TVC, UAB
0.3	2015/07/17	all	First round of contributions included	IRT, TVC, UAB
0.4	2015/08/21	all	Second round of contributions included	IRT, RBB, TVC, UAB
0.5	2015/09/28	all	Fixing details, Executive Summary, included feedback from internal review	IRT, TVC
1.0	2015/09/30	all	Final version	IRT

## Validation

No.	Action	Beneficiary	Date
1	Prepared	IRT	2015/05/15
2	Approved	TVC	2015/09/28
3	Released	UAB	2015/09/30

Disclaimer: The information in this document is subject to change without notice. Company or product names mentioned in this document may be trademarks or registered trademarks of their respective companies.

### All rights reserved.

The document is proprietary of the HBB4ALL consortium members. No copying or distributing, in any form or by any means, is allowed without the prior written agreement of the owner of the property rights.

This document reflects only the authors' view. The European Community is not liable for any use that may be made of the information contained herein.

## Table of Contents

<b>EXECUTIVE SUMMARY .....</b>	<b>6</b>
<b>1. INTRODUCTION.....</b>	<b>9</b>
1.1. PURPOSE OF THE DOCUMENT .....	9
1.2. ACRONYMS AND ABBREVIATIONS .....	9
1.3. DEFINITIONS AND GLOSSARY .....	10
<b>2. SERVICE COMPONENTS .....</b>	<b>13</b>
2.1. CLEAN AUDIO GENERATION (IRT) .....	13
2.1.1. <i>Overview</i> .....	13
2.1.2. <i>Status of completion</i> .....	15
2.1.3. <i>Relation to the sub-pilots</i> .....	15
2.2. ENHANCED AD GENERATION (TVC).....	15
2.2.1. <i>Overview</i> .....	15
2.2.2. <i>Status of completion</i> .....	17
2.2.3. <i>Relation to the sub-pilots</i> .....	17
2.3. MULTIPLE AUDIO ASSET GENERATION (TVC) .....	17
2.3.1. <i>Overview</i> .....	17
2.3.2. <i>Status of completion</i> .....	18
2.3.3. <i>Relation to the sub-pilots</i> .....	18
2.4. HBBTV AUDIO OPTIONS SELECTOR (TVC) .....	18
2.4.1. <i>Overview</i> .....	18
2.4.2. <i>Status of completion</i> .....	19
2.4.3. <i>Relation to the sub-pilots</i> .....	20
2.5. PLAYBACK OF AUDIO CONTENT USING MPEG-DASH (TVC) .....	20
2.5.1. <i>Overview</i> .....	20
2.5.2. <i>Status of completion</i> .....	20
2.5.3. <i>Relation to the sub-pilots</i> .....	22
2.6. HBBTV DEVICE AND VERSION DETECTION (TVC) .....	22
2.6.1. <i>Overview</i> .....	22
2.6.2. <i>Status of completion</i> .....	23
2.6.3. <i>Relation to the sub-pilots</i> .....	23
<b>3. SUB-PILOT INTEGRATION .....</b>	<b>24</b>
3.1. CLEAN AUDIO – SUB-PILOT IN GERMANY (BERLIN-BRANDENBURG) .....	26
3.1.1. <i>Production</i> .....	26
3.1.2. <i>Publishing &amp; Delivery</i> .....	26
3.1.3. <i>Reception &amp; User Application</i> .....	26
3.2. SUB-PILOTS IN SPAIN (CATALONIA).....	27
3.2.1. <i>Production: Clean Audio</i> .....	27
3.2.2. <i>Production: Audio Description</i> .....	27
3.2.3. <i>Production: other languages and language learning</i> .....	27
3.2.4. <i>Publishing &amp; Delivery</i> .....	27
3.2.5. <i>Reception &amp; User Application</i> .....	28
<b>4. SERVICE PILOTS .....</b>	<b>29</b>
4.1. CLEAN AUDIO – SUB-PILOT IN GERMANY (BERLIN-BRANDENBURG) .....	29
4.1.1. <i>Status of work</i> .....	29
4.1.2. <i>Timeframe</i> .....	29
4.1.3. <i>Test users and evaluation approach</i> .....	29
4.1.4. <i>Challenges, problems and potential solutions</i> .....	30

4.2.	CLEAN AUDIO – SUB-PILOT IN SPAIN (CATALONIA) .....	30
4.2.1.	<i>Status of work</i> .....	30
4.2.2.	<i>Timeframe</i> .....	30
4.2.3.	<i>Test users and evaluation approach</i> .....	31
4.2.4.	<i>Challenges, problems and potential solutions</i> .....	31
4.3.	AUDIO DESCRIPTION – SUB-PILOT IN SPAIN (CATALONIA) .....	32
4.3.1.	<i>Status of work</i> .....	32
4.3.2.	<i>Timeframe</i> .....	32
4.3.3.	<i>Test users and evaluation approach</i> .....	32
4.3.4.	<i>Challenges, problems and potential solutions</i> .....	33
4.4.	OTHER LANGUAGES AND LANGUAGE LEARNING – SUB-PILOT IN SPAIN (CATALONIA).....	33
4.4.1.	<i>Status of work</i> .....	33
4.4.2.	<i>Timeframe</i> .....	33
4.4.3.	<i>Test users and evaluation approach</i> .....	34
4.4.4.	<i>Challenges, problems and potential solutions</i> .....	34
4.5.	OVERVIEW OF COMMON CHALLENGES.....	34
<b>5.</b>	<b>COMPLEMENTARY USER TESTS.....</b>	<b>36</b>
5.1.	PRELIMINARY USER TESTS (UNTIL M20).....	36
5.1.1.	<i>Clean Audio</i> .....	36
5.1.2.	<i>Audio Description</i> .....	39
5.2.	SUMMARISED RESULTS AND GUIDELINES / RECOMMENDATIONS.....	42
5.3.	USER TESTS OUTLOOK .....	45
5.3.1.	<i>Clean Audio</i> .....	45
5.3.2.	<i>Audio Description</i> .....	47
5.3.3.	<i>Other languages and language learning</i> .....	50
<b>6.</b>	<b>CONCLUSIONS .....</b>	<b>51</b>
<b>7.</b>	<b>REFERENCES.....</b>	<b>52</b>
<b>8.</b>	<b>ANNEXES.....</b>	<b>53</b>
8.1.	SUB-PILOTS IN SPAIN (CATALONIA) – PUBLISHING AND DELIVERY FOR VoD.....	53
8.2.	SUB-PILOTS IN SPAIN (CATALONIA) – PUBLISHING AND DELIVERY FOR LIVE IP-STREAMING .....	55
8.3.	SUB-PILOTS IN SPAIN (CATALONIA) – RECEPTION AND USER APPLICATION .....	57

## Figures

<b>FIGURE 1.</b> PROCESSING CHAIN OF THE CLEAN AUDIO GENERATOR (BASIC MODE).....	13
<b>FIGURE 2.</b> PROCESSING CHAIN OF THE CLEAN AUDIO GENERATOR (ADVANCED MODE) .....	14
<b>FIGURE 3.</b> AD PRODUCTION WORKFLOW .....	16
<b>FIGURE 4.</b> WORKFLOW ADAPTED TO SUPPORT MULTIPLE AUDIO ASSETS .....	18
<b>FIGURE 5.</b> PLAYBACK AND SWITCHING OF AUDIO COMPONENT USING MPEG-DASH .....	21
<b>FIGURE 6.</b> SYSTEM ARCHITECTURE COMBINING PILOT-B COMPONENTS IN A COMMON VIRTUAL WORKFLOW .....	24
<b>FIGURE 7.</b> RESULTS FOR ONE SPECIFIC TEST CLIP OF THE LISTENING TEST WITH HEARING IMPAIRED PEOPLE THAT DO NOT WEAR A HEARING AID .....	37
<b>FIGURE 8.</b> RESULTS FOR ONE SPECIFIC TEST CLIP OF THE LISTENING TEST WITH HEARING IMPAIRED PEOPLE WEARING A HEARING AID.....	38
<b>FIGURE 9.</b> RESULTS FOR ONE SPECIFIC TEST CLIP OF THE LISTENING TEST WITH HEARING IMPAIRED PEOPLE WEARING A HEARING AID.....	39
<b>FIGURE 10.</b> PUBLISHING/DELIVERY WORKFLOW FOR VOD STREAMS .....	53
<b>FIGURE 11.</b> PUBLISHING/DELIVERY WORKFLOW FOR LIVE STREAMS.....	55
<b>FIGURE 12.</b> RECEPTION AND USER APPLICATION.....	57

## Tables

<b>TABLE 1.</b> OVERVIEW OF COMMON CHALLENGES .....	35
<b>TABLE 2.</b> OVERVIEW OF RECOMMENDATIONS WITH RESPECT TO CLEAN AUDIO .....	44
<b>TABLE 3.</b> OVERVIEW OF RECOMMENDATIONS WITH RESPECT TO AUDIO DESCRIPTION .....	45

## 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, CA), have vision disabilities (Audio Description, AD, 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 reports the progress of Pilot-B within the HBB4ALL project for the first two tasks, in preparation of the operational phase in the project.

Service sub-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 personalised Clean Audio service via IP for a dedicated user group. The sub-pilots in Spain will most likely be carried out as a combined service pilot offering the functionality of all alternative audio services to the end user.

As an overall technical aim, across the mentioned objectives and sub-pilots, 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 broadcast channel. To realise this, the partners had commonly decided to use HbbTV, as this opens up additional possibilities, using the broadband (IP) channel to address also smaller target user groups. For the further technical realisation of the sub-pilots, the partners have been developing and adapting tools, so-called Service Components:

- **Clean Audio generation**, a server-based software that generates an audio signal with improved speech intelligibility. The input signals for the CA generation will be existing content from TV productions (either 5.1 or stereo content). All but one software modules have been realised. Two rounds of lab tests have been carried out; an additional round of user tests is still required specifically for stereo content.
- **Enhanced Audio Description generation**, an extension to TVCs AD workflow that automatically creates an audio mix containing the original audio track and the appropriate AD sentences, timely mixed with the original audio track based on the available metadata. One module is still under development and planned to be ready for use by end of September 2015.
- **Multiple audio asset generation**, an adaptation to TVC’s workflow (in the MAM) to handle multiple audio files with more than the two stereo tracks that were already supported. Also the component prepares the audio assets for publishing as MPEG-DASH assets. This component has been finalised.
- **HbbTV audio options selector**, a module in the TVC HbbTV test application which allows the end user to select, via the user interface, the audio component he is interested in. This component has been finalised.
- **Playback of audio content using MPEG-DASH**, processes multiple audio assets and a video asset to create the MPEG-DASH assets for publishing. The component has been successfully tested in HbbTV testing applications and is thus ready for use in the respective sub-pilots.

- **HbbTV device and version detection**, a module for the TVC HbbTV test application which may be used to adapt the service behaviour (e.g. the version of content delivered) to the device brand or the HbbTV version it supports.

The functionality of these components is explained in this document; most of them are available for use in the respective sub-pilots.

For the technical realisation of the sub-pilots, the partners have been integrating the above mentioned components in dedicated workflows for each sub-pilot. Specifically for the CA sub-pilots, these workflows will contain semi-automatic processing steps in the preparation of the assets for publishing. For the other sub-pilots TVC has realised automated workflows.

The organisational work for the sub-pilots in Pilot-B has progressed and the communication with the user associations was intensified during the preparations. The sub-pilots are scheduled to be carried out within the period of October 2015 – July 2016. Their status is as follows:

- **Clean Audio sub-pilot in Germany (Berlin-Brandenburg)** is currently planned to start in February 2016, with an accompanying user evaluation, to run for approximately six weeks. The planning is still in an early stage and the sub-pilot needs to be further defined, depending on the outcome of the final lab tests.
- **Clean Audio sub-pilot in Spain (Catalonia)** is currently planned to start in January 2016. The definite decisions depend on user tests to be carried out in Catalonia until November 2015.
- **Audio Description sub-pilot in Spain (Catalonia)**, scheduled to start in October 2015. Its final timing depends of the final integration of the automatic AD generation component in TVCs workflow.
- **Other languages and language learning sub-pilot in Spain (Catalonia)**, scheduled to start in November 2015. Some final adjustments to the HbbTV client application have to be done.

For the Catalanian sub-pilots, TVC plans to combine all audio versions into one service offering (e.g. original language, CA, AD and a dubbed version), depending on availability of audio content and the respective rights issues, using the same service provisioning infrastructure.

In preparation of the operational phase, many complementary user tests have been carried out. Tests addressed different aspects of Audio Description, trying alternative versions of AD, the acceptance of synthetic rather than human voices and the use of a Second Screen when accessing AD during a movie. From these tests, dedicated recommendations have been drawn which suggest changes that can improve AD services. Subsequent further user testing, in parallel to the sub-pilots, is planned.

For realisation of Clean Audio, being a novelty as no actual service is on air at the moment, the focus of preliminary user tests was on enhancing the implementation of the CA generator for use in the operational phase. Whereas for 5.1 content good results already had been achieved in lab tests, no conclusive results for using stereo content could be drawn during a second round of lab tests in March 2015. It was concluded that additional tests have to be conducted with a more appropriate evaluation methodology in order to have a good starting position with respect to the creation of audio material for the CA sub-pilots. These additional tests are planned for autumn 2015. Intermediate results and recommendations for a potential CA service are described.

There has been intense cooperation between the partners in Pilot-B. For CA common user tests have been carried out, an intensive exchange of information regarding further CA tests has been done (e.g. regarding search for appropriate content and testing methodology) and will continue into the operational phase. Technical developments for use of multiple audio delivery by means of MPEG-DASH as well as for selection of an audio asset in an HbbTV application have been coordinated and tested.

In summary, the core objectives of the initial tasks of Pilot-B (T4.1 and T4.2), have been largely completed. Since the writing of D4.1, the remaining work for both tasks has progressed well: all service components have been finalised or will be completed shortly; this is also true for the integration of workflows for use in the sub-pilots and the preparation of the various sub-pilots.

Only the work on Clean Audio is somewhat delayed. Additional lab tests for CA with stereo content are required, as the chosen test methodology and test items so far did not allow to conclusively derive generic and stable knowledge. This additional test is planned to be conducted until end of October 2015 and thus will still allow the sub-pilots to be carried out within the foreseen timeframe of the Pilot-B operational phase (in T4.3).

During the following, operational phase (T4.3), the activities of Pilot-B will focus on evaluating the established services in the respective sub-pilots with dedicated user groups.

## 1. Introduction

The Hybrid Broadcast Broadband for All project (HBB4ALL) investigates accessibility services in the hybrid broadcast-broadband TV (HbbTV) environment. One of the most prominent challenges faced by broadcasters is the requirement to add Access Services, in a cost-efficient manner, also to audio-visual content delivered via Internet while remaining consistent with the Access Services available on traditional broadcasts and their respective workflows. An new, additional challenge is the desire to offer viewers the opportunity to customise the Access Services they are using to best meet their personal preferences or needs.

HBB4ALL 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.

The operational phase of the HBB4ALL project (Task X.3 – Operation Phase for all Pilots A to D, running from August 2015 – July 2016) will gather user feedback and assess the acceptance and quality of services in various delivery scenarios implemented using field user tests and also in complimentary qualitative lab tests. A number of different sub-pilots is scheduled to be carried out in the operational phase. In the specific case of Work Package 4 / Pilot-B, novel audio services will be tested and prototypically rolled out in the different HBB4ALL countries in order to tackle the above challenges and receive and evaluate user feedback for later improvement.

### 1.1. Purpose of the document

HBB4ALL deliverable D4.1 – Pilot-B Progress Report (see [2]) provided an overview of the progress of Pilot-B during the first 11 months of the project timeline. The current document gives an update to D4.1, presenting the status of the sub-pilot preparation.

First, the technology and infrastructure, which has been implemented for support of the operational phase, is presented: chapter 2 provides an overview of the prototype service components and their status of completion; chapter 3 documents the integration of components and preparation of the test beds as end-to-end workflows for the sub-pilots. Chapter 3 thus provides a reference point for the reader, how the service components relate to the testbeds / workflows that have been developed for the sub-pilots and how they work “end-to-end”.

Chapter 4 presents the current status of the envisioned sub-pilot preparation with respect to organisational issues, technical and functional setup, timing, envisioned users / user groups and the evaluation approach.

Finally, chapter 5 documents the results of intermediate user tests with respect to the improvement of speech intelligibility by means of Clean Audio. These tests focussed on the improvement / adjustment of the Clean Audio generator to be used in the sub-pilots of Pilot-B. Preliminary user tests for Audio Description and language acquisition also have been carried out, focussing on various aspects, aiming for possible improvement of these services in general. An outlook on further planned user tests accompanying the sub-pilots during the operational phase of Pilot-B concludes this chapter.

### 1.2. 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.

### 1.3. 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.

**Content Delivery Network** (CDN) – Large distributed system of servers deployed in multiple data centres across the Internet. These servers cache and store the content from the Internet content providers that contract this service to enhance the availability and performance of the delivery of content to end-users, while reducing demand on the content provider's own servers.

**Content Management System** (CMS) – A computer application that allows publishing, editing and modifying content, organizing, deleting as well as maintenance from a central interface. Content management systems typically provide procedures to manage workflows in a collaborative environment. (see also MAM).

**ffmpeg** – Complete cross-platform open source software tool for handling and editing multimedia data, both audio and video of various codecs. This solution allows, among other functionalities, the transcoding, multiplexing, demultiplexing, fragmenting, recording and streaming of multimedia files.

**HbbTV** – Hybrid Broadcast Broadband TV is a major pan-European initiative, building on work in the Open IPTV Forum, to formulate standards 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. [For speech perception, the important region is between 250 and 4,000 Hz]. The term 'deaf' is used to describe people with such profound hearing loss that they cannot benefit from amplification, while the term '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.

**Media Asset Management (MAM)** systems are typically software systems to support management tasks and decisions surrounding the ingestion, annotation, cataloguing, storage, retrieval and distribution of audio, video and other media assets.

**Metadata** – Supplementary data about data, in our case information about television programs. E.g this could be 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.

**Media Presentation Description (MPD)** – MPEG-DASH manifest – An XML document which describes segment information for each one of the audio and video components present in an MPEG-DASH stream, and its representations. This segment information contains, amongst others, timing, URL to point to the specific one, duration, audio languages, video width and height, codec and bitrates. This manifest provides enough information for an MPEG-DASH client to be able to access and download the media segments and set an adaptive streaming of the content.

**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.

**System Usability Scale (SUS)** – provides a “quick and dirty”, reliable tool for measuring usability. It consists of a 10 item questionnaire with five response options for respondents; from Strongly agree to Strongly disagree. Originally created by John Brooke in 1986, it allows you to evaluate a wide variety of products and services, including hardware, software, mobile devices, websites and applications<sup>1</sup>.

**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.

---

<sup>1</sup> <http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>

## 2. Service Components

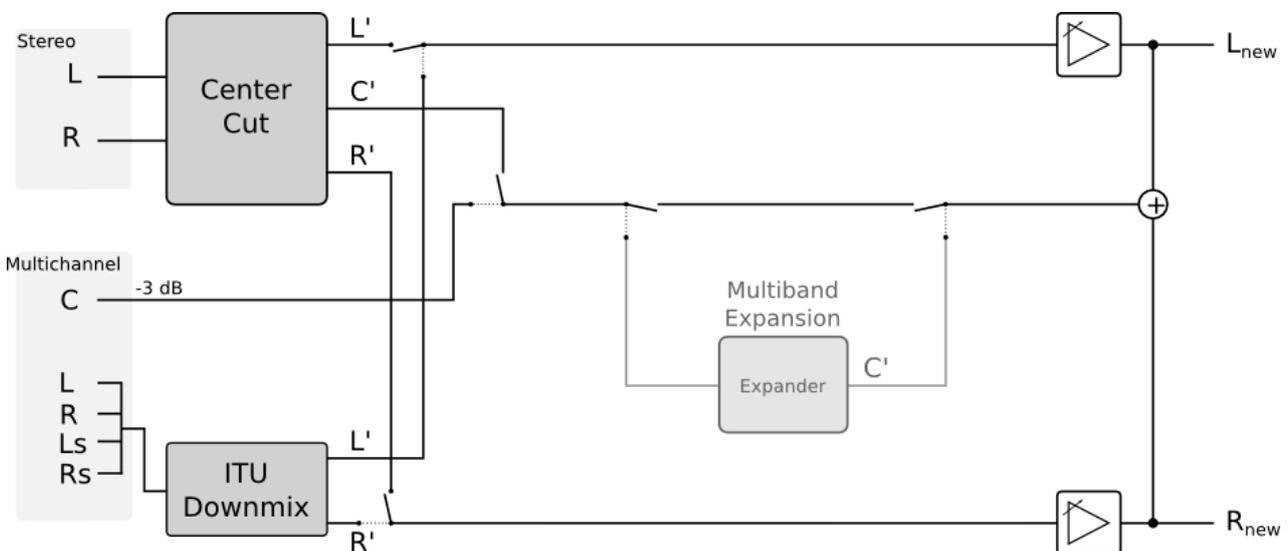
This chapter provides an update with respect to D4.1 [2], giving an overview of the Service Components which are being developed and/or adapted for use in the sub-pilots of Pilot-B. Specifically, for each component it is described what has been completed so far and in which sub-pilots it will be used (see also chapter 3 on sub-pilot integration).

### 2.1. Clean Audio generation (IRT)

#### 2.1.1. Overview

Due to the limitations of the HbbTV 2.0 standard and restrictions on the audio production side, it was decided to implement the Clean Audio (CA) sub-pilot as server-based approach instead of a client-based approach, which means that pre-produced content will be transmitted to the user. The input signals for the CA generation will be existing content from TV productions (either 5.1 or stereo content).

The main requirement for the aimed software tool ('Clean Audio generator') is an automatic generation of CA versions. **Figure 1** and **Figure 2** show the processing chain of the component for both modes basic and advanced. The basic mode will mainly only enhance the relation of speech to the rest of the signal.



**Figure 1.** 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 [3] is applied onto 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. As a next step, the volume of the modified L' and R' channels will be turned down by applying a negative gain factor (e.g. -9 dB). Finally, the newly generated Center signal will be added to the reduced L' and R' channels. The resulting stereo signal (containing L<sub>new</sub> and R<sub>new</sub>.) will be transmitted to the user.

For 5.1 content, there is no special speech extraction step required, since the Center signal (which in most cases only contains speech) is already available.  $L'$  and  $R'$  are generated by applying an ITU downmix [4] on the L, R, Left Surround and Right Surround channels of the 5.1 feed. Eventually, the volume of  $L'$  and  $R'$  is reduced by a negative gain factor and the Center channel will be added with original volume. The result is again a stereo signal containing  $L_{new}$  and  $R_{new}$ .

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 will be applied to the Center. Firstly, if the dynamic (relation of slightest and loudest level) of the audio signal in certain (speech relevant) frequency sub-bands is too small due to heavy compression during the audio production, it will be enhanced by applying an expander on the speech frequencies (“Multiband Expansion”). Expanding means that potential silent (noise) signal parts will be even further reduced (making quiet sounds even quieter).

During a second additional step, the signal is filtered with a bandpass and bandstop filter (“Frequency Processing”). 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. The bandpass (currently 1.6 kHz – 8 kHz) will be added with its 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.

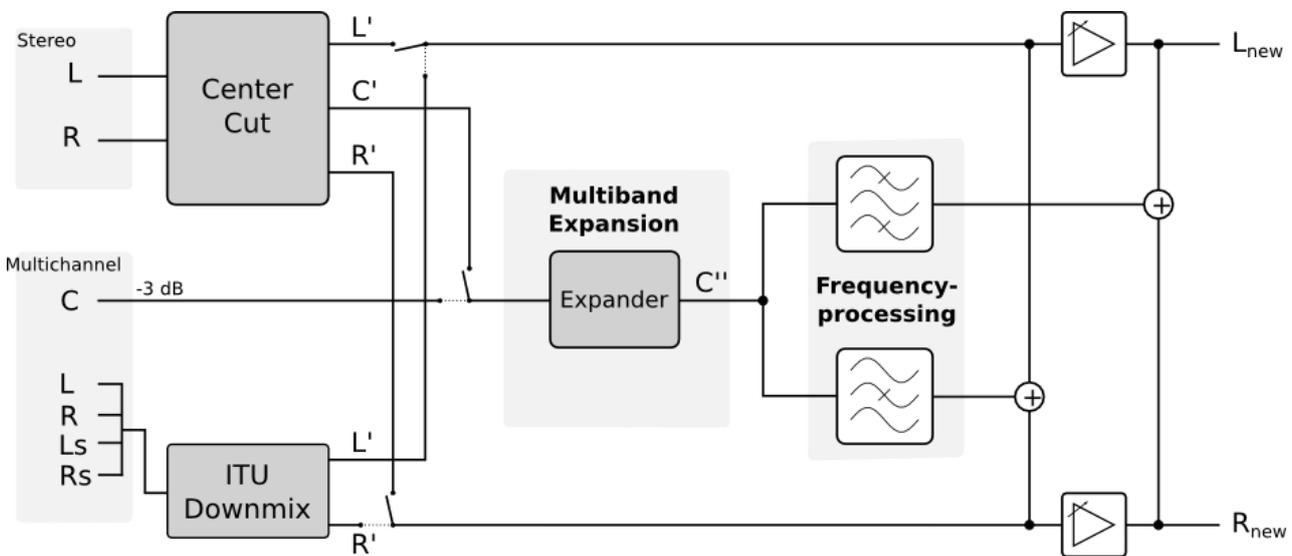


Figure 2. Processing chain of the Clean Audio generator (advanced mode)

### ***2.1.2. Status of completion***

The CA generator has been implemented in Python as an integrated software solution for an automatic usage on different platforms and operating systems. The following modules have been completed so far:

- CenterCut algorithm
- ITU downmix
- Frequency processing
- Input / output of the audio signals

Not yet integrated is the Expander module. This last step should be completed within the running time of Task 4.2. Furthermore, some extensions and improvements for the software for a more adaptive processing of the audio signals might be beneficial.

The results of the listening tests and the feedback of the test subjects (see section 5.1.1) lead to the conclusion that the level of enhancement (improvement of the speech intelligibility) highly depends on the original signal in terms of mix and capturing situation.

### ***2.1.3. Relation to the sub-pilots***

The Clean Audio generator will be used for the generation of additional audio tracks for the German sub-pilot with RBB (see sections 3.1 and 4.1). Moreover, it is planned to use it to create content for a Catalan CA sub-pilot by TVC/UAB (see sections 3.2.1 and 4.2) as well as for CA-related tests planned by UAB (see section 5.3.1.2).

## **2.2. Enhanced AD generation (TVC)**

In the production workflow for live broadcast TVC has an automatic Audio Description (AD) generator based on Text-To-Speech (TTS) already in use. This is used in some informative programs like ‘TN’ (the TVC TV news program), to automatically convert subtitles to AD. Originally, it was planned to adapt this for HBB4ALL, but the tool is very specific for the live broadcast case, and for some specific program types.

Therefore, it was decided to develop an AD generation component for HBB4ALL, which has a larger added value for the TVC workflow and which can be applied generically for any type of program. This component is described in the following sections.

### ***2.2.1. Overview***

For the enhancing of the AD generation as an automatic process included in the production workflow TVC has been developing three different sub-components with specific functionalities, see **Figure 3**.

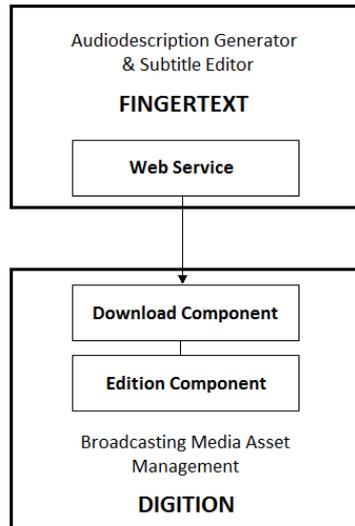


Figure 3. AD production workflow

### HTTP Web Service to provide audio description contents

The FingerText system, operated via web interface and installed at TVC, generates and stores a set of files with Audio Description messages and associated metadata in a database. For HBB4ALL an HTTP Web Service has been developed that allows downloading these contents related to a video Asset, after authentication. The service call requires the Asset code identifier, which we name “codi Preparat Per Difusió” (prepared for diffusion – PPD code). The result of the download is a compressed file in ZIP format which contains:

- MP3 files, each one corresponding to an Audio Description phrase
- An XML file describing for each MP3:
  - The position (TimeCode) when the file has to be played
  - The level of attenuation that has to be applied to the original audio soundtrack in the background when the MP3 is played
  - A stamp indicating that the attenuation has to be kept after the file finishes playing. This is necessary to avoid ups and downs in the audio volume when there are several phrases that are played within a short time interval between them

### Download Component

This component makes a call to the HTTP web service explained before and downloads the ZIP file associated with the PPD code to unzip all MP3 and XML files.

## Edition Component

The Edition Component reads the WAV file containing the original or Catalan audio for a specific asset. At the same time it takes the AD MP3 files downloaded by means of the Download Component and processes a new WAV file with the original WAV audio attenuated and mixed with all AD MP3 files inserted at time stamp positions as described in the XML file.

### *2.2.2. Status of completion*

The first component described before, the HTTP web service, is completely developed and correctly working to date. The Download Component is currently being developed and has not yet been finished. The audio editor is in a study phase, it was planned to be developed using ffmpeg or Microsoft Directshow Editing Services (DES) libraries, and TVC finally decided to use DES. It is planned to have the Edition Component ready for use by end September 2015.

When all sub-components mentioned are finalised, they will be integrated in TVCs Export Engine. The Export Engine is the software in use by TVC to encode MP4 video files (and associated metadata) to be published to the web and the HbbTV services. Basically it executes the encode requests which are generated by different workflows. In case this export engine receives an encoding request for AD content generation, first a WAV file with the audio-description mix will be generated (using the components discussed in 2.2.1), then the encoding to a MP4 file will be done from the original video file and the new WAV file.

### *2.2.3. Relation to the sub-pilots*

This service component will be used for the content generation for the AD Catalan sub-pilot, to offer the AD content with the original audios (see sections 3.2.2 and 4.3).

## **2.3. Multiple audio asset generation (TVC)**

### *2.3.1. Overview*

At the start of HBB4ALL, TVC's workflow was already able to work with video assets containing two stereo audio tracks. TVC has adapted its workflow to handle multiple audio files with more audios than the two stereo tracks that were already supported (see **Figure 4**). The generation of multiple audio assets for a single video asset is primarily needed for the transport and delivery of different audio languages, AD, CA etc. in the same file, using an MPEG-DASH (Dynamic Adaptive Streaming over HTTP) based streaming solution.

TVC developed its own Broadcasting Media Asset Management (MAM) system, called Digion, and also developed the IP Publishing MAM system called Deliverity. Both systems are continuously evolving as MAM requirements grow. The MPEG-DASH content generation, from the assets present on Digion, takes place based within this TVC workflow. Using Deliverity, the contents that TVC wants to test or offer as MPEG-DASH streams are exported to TVC's MPEG-DASH generation/delivery server. That is where these contents are processed and the required MPEG-DASH segments and Media Presentation Description (MPD) files are generated, using a specific tool or MPEG-DASH server solution.

Apart from the workflow adaptation, a set of processes was standardised to be followed by the studios to edit the different output audios with the correct formats and to include all the tracks with the video in a single file.

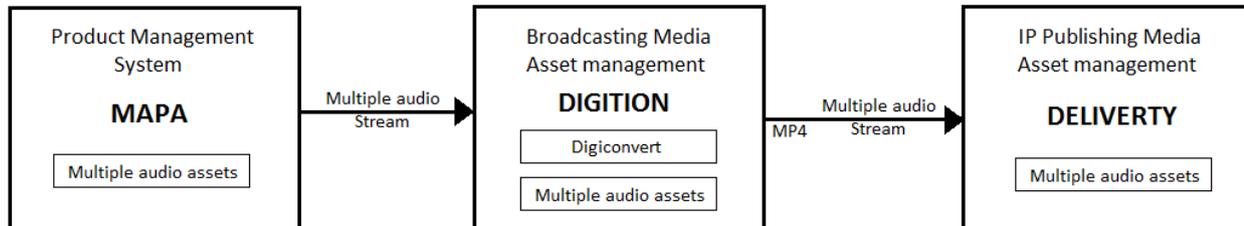


Figure 4. Workflow adapted to support multiple audio assets

### 2.3.2. Status of completion

The Media Asset Management (MAM) system used by TVC (Digition, see **Figure 4**), which stores the contents to be published, has been adapted to store files containing a video and multiple audio files. The video is encapsulated in AVI files and can be encoded with different codecs (DV, DVCPRO, DVCPRO50 or DVCPROHD). The audios are encapsulated in WAV files. The multiple audio languages assets stored in Digition can have different formats to comply with the inclusion of original language and Catalan audios in the same file (Dual mono, Dual stereo, Dual 5.1 and multi-audio).

The publishing of the contents by TVC is always based on using MP4 files with stereo audio. Because of that, a component called Digiconvert has been developed (see **Figure 4**), which allows encoding the Digition files in MP4 format, processing the different formats of audio commented before and thus preparing the audio assets for publishing as MPEG-DASH in the next step. This component has been finished but at the time of writing has not yet been integrated in TVC's Export Engine.

### 2.3.3. Relation to the sub-pilots

As described before, TVC first envisioned this component to be used only for the 'another languages and language learning' Catalan sub-pilot. In the course of the project it has been decided to use the component in all three Catalan sub-pilots (see section 3.2.4). The CA and AD audio tracks will also be handled as multiple audio assets in the respective sub-pilot to deliver the respective audio service in addition to the original one.

## 2.4. HbbTV audio options selector (TVC)

### 2.4.1. Overview

TVC has developed an audio selector which is accessible via the player bar in the user interface of the HbbTV application, to allow users to switch between the different audio options available. This audio selector is Javascript based and uses the API's (Application Programming Interface) functions for the selection of specific A/V components for playback, as defined in the Open IPTV Forum (OIPF) Specification [6]. This functionality can run on devices supporting HbbTV 1.5 or higher using MPEG-DASH, and works for Video On Demand as well as Live streams.

The audio options selector Graphical User Interface (GUI) consists of a dropdown menu which contains all different audio options available in the stream and allows the selection of any of them to be played. To

correctly identify the different options TVC has used the language property of the audio components in MPEG-DASH, according to the ISO 639-2 standard [7] wherever possible. For the options that do not have a specific language code in that ISO standard, like Clean Audio or Audio Description, there is a range of codes reserved for local use that we could profit of. The dropdown menu is located on the player bar, and when one audio is selected it shows the language descriptor of the audio currently being played. When the user navigates to the component and clicks on it, it displays the complete set of options available (see Annex 8.3, **Figure 12**).

This selector has been developed focusing on switching between Audio Components present in the same MPEG-DASH stream. Also, for the cases where the use of MPEG-DASH is not possible, it can be adapted to work with HTTP download on separate MP4 files, requesting the alternate file and seeking the video and audio content to the current playback position.

### ***2.4.2. Status of completion***

At the time of writing the HbbTV audio options selector for MPEG-DASH multiple audios delivery has been developed in testing applications and it is not yet integrated in any public TVC service. For that, the characteristics of the selector have been manually set for testing its functionality for different streams and amount of audios. A future development could include a fully automated selector, regarding automatic adaptation of the selector's audios list to the number of audios of each content being played, or identifying the brand of the TV client and prepare the selector to work for selecting separate audios of the same video asset knowing if they correctly support the MPEG-DASH multiple audio and component selection or not.

For the selection of a certain audio component out of those available in the MPEG-DASH stream, various mechanisms have been tested. In the best case scenario, this selector is capable of setting the different available audio options from the information present in the MPEG-DASH Media Presentation Description (MPD) file, identifying and separating them by their language, mapping their language with the corresponding position in the Audio Components Collection of the MPEG-DASH stream. With this, the order they appear in the selector's list is the same order as they have in the collection of audios of the stream. Then, when another audio is selected by the user, the selector finds its position in the selector's list by searching its language and sets as the next Audio Component to be played the one located in that position. In that case the player and the audio selector would not need any previous information about the audios of the content being played: all that it needs could be obtained from the MPD and it could be implemented for the most number of audios and prepared to work from only one audio to this highest number of different audios.

In a previous version of the selector, when the audio changed, the selector only returned the language identifier instead of searching its position in the selector's options. Then, the searching for the audio position in the list was done by checking every Audio Components and its language until finding the one selected by the user. In the way explained before, the audios list of the selector is only an array of strings, containing each language at each position, and it's more easily parsed than the Collection of Components.

The above described first case depends on the correct HbbTV client implementation. In some cases the language of the audio components is not correctly retrieved by the TV client; tests carried out by both TVC and IRT with HbbTV devices of various manufacturers have confirmed this. To address this problem, some additional information about the audio languages and its relative position in the Audio Components list of the MPEG-DASH stream must be known, and used by the selector if it cannot retrieve a language or it is undefined. Taking into account that the streams will be played by HbbTV applications managed by the broadcasters we can consider that the information about the language of the different audios available will be

known and carried among other metadata (titles, subtitles, descriptions, durations, URL of the video content itself...) without entailing significant additional effort.

Another open issue which was found during the tests is that some HbbTV devices cannot navigate properly inside TVCs audio component selector. While the navigation through the player bar buttons is correct in all the tested devices, in some of them – when the dropdown menu is displayed – in the navigation among the different audio options list one position is skipped. When having more than two audios, if the user tries to go up or down in the audio options list the focus passes to the option two positions up or two positions down, skipping the very next one.

### ***2.4.3. Relation to the sub-pilots***

The audio components selector will be highly probably included in the applications used for all the Catalan sub-pilots of the Pilot-B related to the three different audio options (Clean Audio (CA), Audio Description (AD), and other language audios, see section 3.2.5).

The most effective case is including all this three targets in the same HbbTV application, and with TVCs audio selector this can be realised. The audio options could be prepared to include the different generated audio streams. If content with two different audio languages is selected (for example: original version in English, and dubbed in Catalan) and for the same content a Clean Audio version of any of the languages and an Audio Description is provided, all audio options can be offered to the end users, who may use the audio component selector functionality to choose the appropriate audio track in the respective sub-pilot (see Annex 8.3, **Figure 12**).

## **2.5. Playback of audio content using MPEG-DASH (TVC)**

### ***2.5.1. Overview***

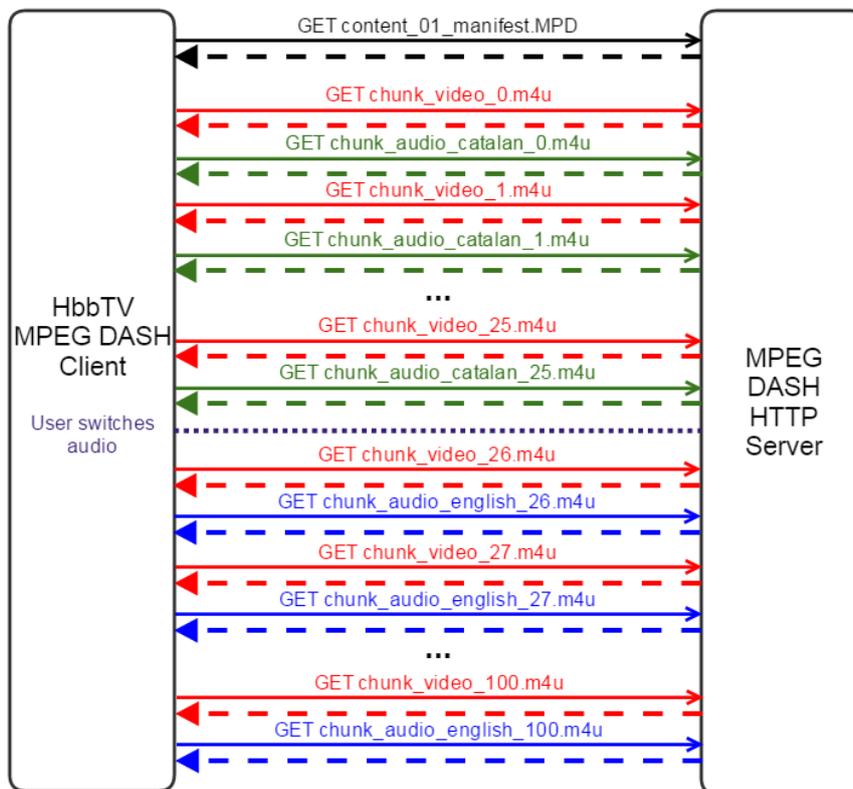
To serve and play multiple audio contents the additional audio streams, for Clean Audio, Audio Description or different languages, are included as separated audio components in the same MPEG-DASH asset. For example, in the simplest scenario, contents with one video and two different audios, the MPD manifest contains three MPEG-DASH adaptation sets: one for the video component, one for the first audio component, and one for the other audio component. The end user, by means of the HbbTV player on the end user device, can select the audio stream he wants to be played among all the audio components presented in the MPD, as explained in section 2.4.

### ***2.5.2. Status of completion***

The generation of MPEG-DASH assets with multiple audio components, their playback and the selection among them has been tested in HbbTV testing applications, with satisfactory results; the feature has not yet been integrated in any public TVC service. In the HBB4ALL Catalan sub-pilots an HbbTV application will be used, based on the testing applications, which is independent of the public TVC service. The integration of the component in an existing TVC service will depend on the outcome of the pilot phase. With different solutions for the Video On Demand or the Live case, the MPD and segments can be generated for streaming MPEG-DASH contents with more than one audio component available. Videos with more than two different audios have been tested and worked correctly in some of the HbbTV 1.5 devices.

To separate the different audios in independent adaptation sets the language attribute of the audios has been used (the use of this attribute has been agreed upon between TVC and IRT as the most efficient implementation for the HbbTV application). With the tools used for the segmentation and MPD generation, if the audio components did not have different “language” attributes only one of the audios was used and presented in the MPD, and the others were ignored. Also, in all cases tested, each audio component has only one representation, one quality and one bit-rate, which means they are not adaptive in that sense. Checking the multiple audio and audio selection behavior with adaptive audios with several representations in bandwidth constrained environments could be a possible future testing to be developed.

When the playback starts the video component and one of the available audio components is selected, the HbbTV client begins to request the corresponding video and audio chunks (see **Figure 5**). Then, when the user changes to another audio component, this component is selected and the client requests the corresponding chunks. The request for video chunks continues while the requests for audio chunks will stop and restart (replay) with the new audio chunks.



**Figure 5.** Playback and switching of audio component using MPEG-DASH

During the tests, two issues have been detected related to playback when the audio component being played is changed:

- 1) When a new audio is chosen by the user the audio switching is not immediate: the time between the selection and the playback of the audio in the device can go from milliseconds to the duration of the MPEG-DASH chunks of the content, depending on the HbbTV device type and brand. One open issue is that, while this is happening, the video playback stops at the device and only continues when the chunks of the newly selected audio are available for playback – in the meantime a black screen is

shown in the player. On other devices, the behavior may be slightly different, instead of a black screen the player shows the last video frame being played at the moment the user selected the new audio, like the video playback was paused instead of stopped.

- 2) When the A/V playback restarts with the newly selected audio the playback position jumps forward or backward in time. The time to restart play with the new audio can be as long as the chunk duration. If the user changes the audio component with the playback while it is just in the middle of a chunk, the same chunk of video is repeated from the beginning with the new audio chunk corresponding to the same instant of content that was previously being played. The same way, if the audio change is done just at the beginning of a chunk the jump is only of few milliseconds, and if the change is started at the end of one chunk the content jumps forward some milliseconds, too. However, the behavior is not deterministic: sometimes there is no temporary jump, in other cases the same video chunk is played again from the beginning with the new audio, or the remainder of the current chunk is skipped and the next one is played.

### ***2.5.3. Relation to the sub-pilots***

As explained in section 2.4.3, each of the Catalan sub-pilots in Pilot-B will make use of at least two audio components (original language + CA / AD / another language), and thus generating MPEG-DASH streams containing multiple audio components in the same asset would also be needed in all of them (see sections 3.2.4 and 3.2.5). If possible, the various audios will therefore be combined into one MPEG-DASH stream. The extent to which this is possible depends on the availability of audio content and the clearing of rights' issues.

In case MPEG-DASH is used in the Berlin-Brandenburg sub-pilot, the same approach for audio asset preparation and audio playback/selection would be used<sup>2</sup>. IRT and TVC have been in close contact and exchanging testing results with respect to the delivery and playback of multiple audio components in a single MPEG-DASH stream. The generation of MPEG-DASH assets and MPD file as well as the component selection implementation in the end user application is shared amongst all sub-pilots in Pilot-B.

## **2.6.HbbTV device and version detection (TVC)**

### ***2.6.1. Overview***

Detecting the concrete device and the supported HbbTV version is necessary to ensure that the features used in the running HbbTV application are compatible with this specific device and will work correctly. This detection is done by retrieving the device's user-agent.

If some functionalities require a higher HbbTV version than the one implemented on the device these functionalities will not work and potentially its usage could make the application crash. In that case, the application should know it and adapt to use alternative ways to execute the same operations or try to avoid using them, if possible. And, if that's not possible, the HbbTV version detection could be even useful to determine the cause of problems or bad behaviours in the devices.

In the same way, the device detection can determine a specific TV brand, or device vendor, which could have a wrong behaviour or known open issues that affect all devices of this brand similarly. With that,

---

<sup>2</sup> The use of MPEG-DASH for the Berlin-Brandenburg sub-pilot is currently not foreseen; this must be seen as a provisional statement.

knowing the device which is running the HbbTV application some vendor-specific corrections or adaptations of its functions could be added to the application, as far as possible.

### ***2.6.2. Status of completion***

A JavaScript component has been developed which takes the user-agent of an HbbTV client device and from that detects the device and the HbbTV version implemented on it.

The devices available at TVC for testing all have the same HbbTV version (HbbTV 1.5), and in most of the cases multiple audios are not yet correctly supported. In some devices, the playback cannot even start if the MPEG-DASH stream includes more than one audio component, or it can start but the audio switching by means of the selector (explained in section 2.4) does not change the audio being played or crashes the content playback. The models that behave properly are from two different brands, and among these the user-agent detection was not actually required because all of them can execute correctly the same functionalities, without added or alternative functions to run based on the HbbTV device and version.

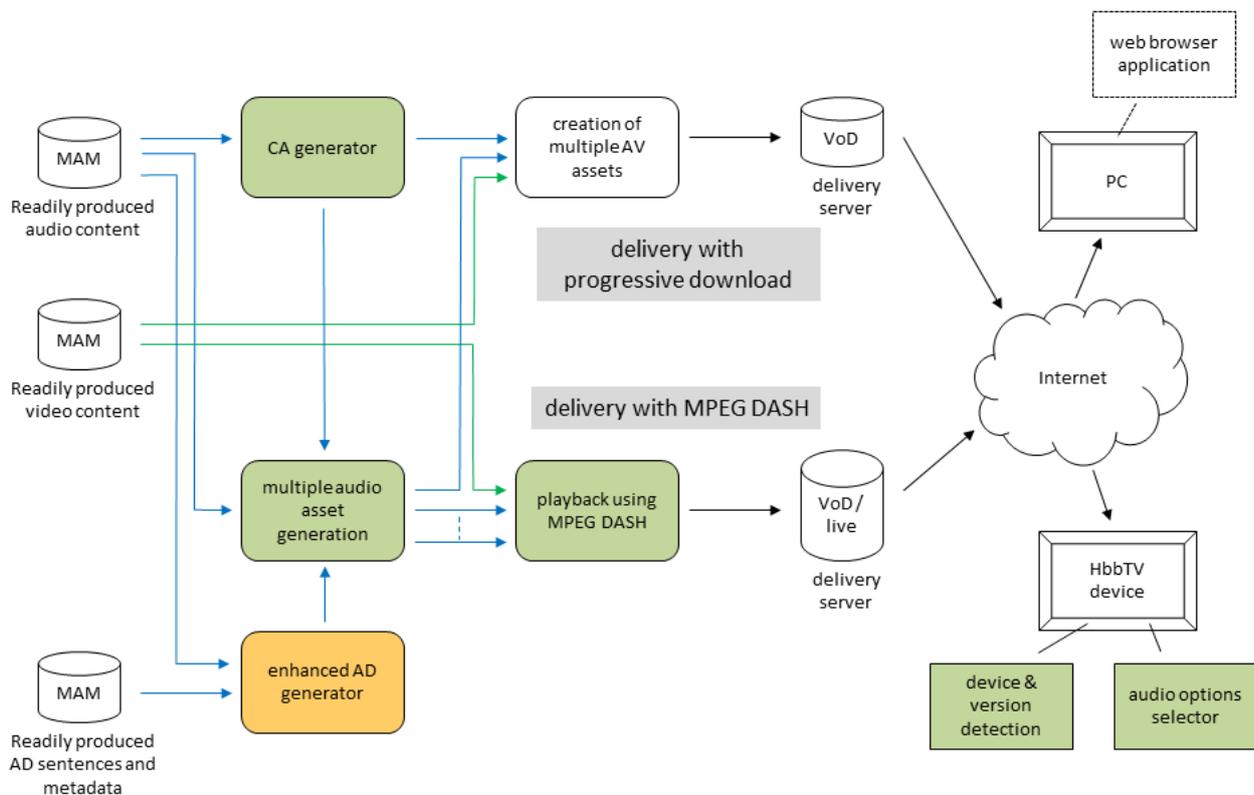
In an application for testing the multiple audio functionalities within the Catalan sub-pilots, or for integration of this feature in an existing TVC service, the detection component will be added. Even without having different solutions depending on the device or HbbTV version, this detection would allow making the multiple audio features only available for the TV models and versions involved in the sub-pilots (which have been successfully tested before).

### ***2.6.3. Relation to the sub-pilots***

This device and version detection component will be used for all Catalan sub-pilots involved in Pilot-B (see section 3.2.5). With this component the sub-pilot can focus on offering the various functionalities (audio selection, MPEG-DASH based streaming), for a specific HbbTV device, as could be the particular TV model(s) tested successfully on beforehand. In this case the respective HbbTV application would be made only visible or displayed for those users that have the desired TV model and not accessible for the rest.

### 3. Sub-Pilot integration

This chapter for each sub-pilot provides an overview of the end-to-end workflow planned for use in the operational phase of Pilot-B, with respect to the following parts in the chain: Authoring and Production, Contribution and Publishing as well as User application and Rendering. For each sub-pilot the Service Components which are being used are referenced (see chapter 2 for detailed descriptions). The joint technical activities between the partners are also summarised in D2.3.2 [1].



**Figure 6.** System architecture combining Pilot-B components in a common virtual workflow

The general system architecture overview shown in **Figure 6** illustrates how an audio workflow could look like if all components would be combined. As not all components are used in each sub-pilot, moreover, not all components will be integrated in actual audio workflows and some represent semi-manual processing steps, this does not represent a real-life situation but merely shows a potential “full integration” to illustrate the potential of the combined output generated by the partners in Pilot-B and to further clarify the relationships between the components described in chapter 2. The components that are being developed in HBB4ALL are highlighted; those that are marked green are used in multiple sub-pilots.

All Pilot-B components are built upon the premise that readily produced content will be reused. This would allow an easier and cost efficient integration in existing workflows. In practice this means that the components may handle 5.1 and stereo audio signals at their input. In **Figure 6** this is illustrated on the left hand side by the Media Access Management (MAM) systems. Following workflow steps can be distinguished:

### Creation of sub-pilot content:

- *Clean Audio*: The *CA generator* processes 5.1 or stereo content to create one or multiple CA versions of the input signal. For use in the sub-pilots this component will probably run semi-automatically; an operator is needed to select the input signals and to transfer the output signals to the publishing part of the workflow.
- *Audio Description*: The *enhanced AD generator* automatically creates an audio mix containing the original audio track and the appropriate AD sentences (timely mixed with the original audio track based on the available metadata, this component will only be used in the AD sub-pilot in Catalonia).
- *Multiple Languages*: The original sound track as well as (a) dubbed version(s) (or any other language versions) are readily available from the audio MAM and are further processed “as is”.

### Asset publishing and delivery:

- *Progressive download*: For a Video on Demand (VoD) service using progressive download, for each audio version a separate AV asset (containing the video and one audio track) would be provided. This step may be done manually in HBB4ALL, depending on the number of assets that need to be processed for the sub-pilots. Specifically this may be required for the CA sub-pilot in Berlin-Brandenburg, where it is planned to provide multiple CA versions to the end user.
- *MPEG-DASH*: Before creating MPEG-DASH assets, the audio signals are converted into the appropriate format; for this purpose the *multiple audio asset generation* component has been developed. It processes any audio signal in Pilot-B and generates a version in MP4 format for use in the MPEG-DASH delivery.

The *playback using MPEG-DASH* component processes a video asset with multiple audio assets to create the MPEG-DASH assets for publishing. The thusly realised assets will be hosted on existing server platforms for access by the end user and distributed via Internet.

### Reception / end user platforms

- *PC*: Wherever PCs are used in the sub-pilots, a web browser application will be developed to allow end users access the content in the service offered. This does not require a full blown application. It is planned to use PC access for the users taking part in the sub-pilot in Berlin-Brandenburg.
- *HbbTV device*: For the use of HbbTV devices in various sub-pilots, several specific components have been developed as part of the appropriate HbbTV application:
  - o The *device & version detection component* may be used to adapt the service behaviour (e.g. the version of content delivered) to the device brand or the HbbTV version it supports.
  - o The *audio options selector component* allows the end user to select, in the HbbTV application user interface, the audio component he is interested in.
  - o The device-native MPEG-DASH player is used to retrieve and render the MPEG-DASH streams.

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

The way how the CA sub-pilot in Berlin-Brandenburg will be realized strongly depends on the final results of the outcome of additional lab tests in autumn 2015 (see section 5.3.1.1) and the availability of content (specifically 5.1 or stereo). Therefore, the planning for the German sub-pilot is still in an early stage.

#### 3.1.1. Production

It is planned to produce the Clean Audio material to be published in this sub-pilot in four steps:

1. Video processing:  
RBB will select a number of video items from its Content Management System (CMS)<sup>3</sup>, where all video content is managed for online and HbbTV publishing. These video items will then be copied and stored locally for further processing.
2. Clean Audio generation:  
The selected video items will be processed with the help of the *Clean Audio generation component*, see section 2.1, either on site at RBB or by IRT, depending on e.g. the amount of content to be processed and the ease of on-site installation.
3. Re-transcoding:  
The processed video items then will be re-transcoded in the pre-defined and ARD-specific video (incl. audio) formats.
4. Upload:  
The Clean Audio versions of the selected video items will then be uploaded to the CMS as new content items, labelled as Clean Audio material. RBB will also consider alternative content storage options for dedicated use in the sub-pilot, in case the usage of RBB's public CMS system is not possible, e.g. due to senior management decisions.

#### 3.1.2. Publishing & Delivery

For publishing and delivery RBB is considering two possibilities: the usage of RBB's standard online CMS or a simple VoD web server, dedicated for this sub-pilot.

#### 3.1.3. Reception & User Application

However the publishing and the delivery would be implemented, RBB is evaluating two ways for the sub-pilot users to access the content as a service. It is possible to set up a dedicated web site for the Clean Audio material to be received, holding a simple video player (accessible by means of PC+webbrowser). A different approach could be to make use of the HbbTV application used for the German signer sub-pilot as planned for Pilot-D (WP6), by adapting the set-up of the application itself, and configuring it for a simple playout of a number of video items containing Clean Audio. The application will be developed to allow for such a re-configuration.

---

<sup>3</sup> Although there may be functional differences between a CMS and a MAM, they are comparable with respect to the architecture as shown in **Figure 6**.

## 3.2.Sub-Pilots in Spain (Catalonia)

With respect to integration, the foreseen sub-pilots in Catalonia largely rely on the same infrastructure. Therefore, they are combined in this section: first, for each sub-pilot the production of the required assets is described separately (sections 3.2.1, 3.2.2 and 3.2.3); then the common Publishing and Delivery as well as Reception & User Application is described (sections 3.2.4 and 3.2.5).

### 3.2.1. Production: Clean Audio

For the production of CA, TVC plans to test the IRT *Clean Audio generator component* (see section 2.1) and evaluate the options of integrating it in TVCs workflow. Three aspects will be considered: if it is possible to integrate the CA generator without problems, the feedback during the Catalan sub-pilot from the target users, and finally the licensing issues that may be involved for TVC for using this component.

Most likely, IRT will process CA versions of the audio source material for use in user tests by UAB and in the Catalan sub-pilot, at least to create early test material which can be ingested into TVCs workflow (to include CA in the corresponding media asset content file, and generate the MPD and MPEG-DASH chunks). In parallel the possibilities for an on-site installation of the *Clean Audio generator* at TVC will be evaluated.

### 3.2.2. Production: Audio Description

The production of the AD files for the sub-pilot testing will be carried out by TVCs *AD generation component*, as explained in section 2.2, integrated in TVCs production workflow. As mentioned in section 2.2, this component will take as input a file with the original audio, without AD, and the MP3 AD files and the XML. This component will mix them and generate as output an audio file with the desired audio description.

### 3.2.3. Production: other languages and language learning

The production of the other languages, as original language for language learning, will be carried out in TVCs dubbing studios. An audio file in another (not Catalan) language will be dubbed to Catalan following the needed format requirements to output each audio. Then, the produced media asset content file will include both audios and will be ingested by TVCs general workflow, adapted to work with the multiple audio files (see [2], sections 3.2.2 and 3.2.4).

### 3.2.4. Publishing & Delivery

The Publishing & Delivery part of the workflow is the same for all three Catalan sub-pilots, as they all deliver an additional audio service to the end user from which a user can select whichever he requires. All three types of audio services evaluated in the sub-pilots (CA, AD and other languages) will be integrated into the same HbbTV application for the Catalan sub-pilots.

This part of the workflow takes as input the output of the production workflows described in sections 3.2.1, 3.2.2 and 3.2.3, or all of them, starting its work from a media asset content file with multiple audio components (original audio in a specific language (original version preferably, not dubbed audio) plus CA or/and AD or/and an audio in another language (probably a dubbed one)) in MP4 format (generated by the *Multiple Audio Asset Generation component*, see section 2.3). For all the additional audio services first the video and the audio components need to be encoded with the required characteristics and output formats. The

required coding depends on the MPEG-DASH server or tool solution used in the sub-pilots (as described in the *Playback of audio content using MPEG-DASH component*, section 2.5), which can require specific characteristics for video or audio contents at its input to work properly. Ideally, in the future, if a specific server solution is chosen for long/mid-term and the additional audio services are integrated in the TVC workflow, and respecting the outputs formats required by TVC's broadcasting service, the coding could be set directly at the output of the production components.

TVC uses different publishing and delivery tools for the VoD and Live IP-streaming case. On both cases the primary objective of these tools is the same: generate adequate MPD and segment files of the separate A/V components as output, placing them in specific directories. In turn, these directories are under the content directory of an HTTP server, thus the manifest and chunks which are part the MPEG-DASH stream are accessible and available for delivery to HbbTV devices, via public URLs. These URLs point to TVCs public server through an external Content Delivery Network (CDN) service.

The planned VoD implementation for publishing and delivery is based on MP4box; details can be found in Annex 8.1. The live IP-streaming case is based on Unified Streaming Platform, see Annex 8.2.

### ***3.2.5. Reception & User Application***

The Reception & User Application part of the workflow is the same for all three Catalan sub-pilots as all additional audio services delivered to the end user are integrated in a single HbbTV application in which a user can select whichever audio component he requires by means of the *HbbTV audio options selector* component (see section 2.4).

This part of the workflow, for the reception of the multiple audio contents, is implemented on the client side. TVC has developed an HbbTV application to be launched in the user device, a SmartTV with Internet connection, to request the MPEG-DASH streams, play them and test the additional audio services involved in the sub-pilots. This application can request for playback any of the contents available in TVCs delivery Servers, VoD or Live, via a URL pointing to the specific MPD of the desired MPEG-DASH content file. As soon as the HbbTV application gets the MPD it starts requesting the A/V chunks of the assets corresponding to the user's selection. If all works fine, these chunks arrive consecutively to the HbbTV device and are played in the MPEG-DASH player of the application.

The sub-pilot will probably be focused to users with a concrete TV brand and model the HbbTV device and the version detection component will also be included in the application, allowing to display the HbbTV application only to those users that have the desired TV model and not accessible for the rest (the *HbbTV device and version detection* component is used for this means, see section 2.6). This specific device would have been previously tested by TVC to ensure its correct behaviour with our application and functionalities.

Details of the HbbTV application can be found in Annex 8.3.

## 4. Service pilots

The HBB4ALL sub-pilots are scheduled to be carried out within the period of August 2015 – July 2016. This section presents the status of each sub-pilot preparation with respect to organisational preparation, technical and functional setup, timing, envisioned users / user groups and the evaluation approach.

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

#### 4.1.1. Status of work

As already mentioned in section 3.1, the way how the CA sub-pilot in Berlin-Brandenburg will be realized strongly depends on the outcome of additional lab tests until October 2015. Whereas for 5.1 content good test results already had been achieved (see [2], chapter 7.1.1), no conclusive results for using stereo content could be drawn during a second round of lab tests (see section 5.1.1). Therefore it is deemed necessary to carry out additional tests with a new testing methodology (see section 5.3.1.1). These lab tests originally were not part of the work plan and thus the start of this sub-pilot will be delayed. Therefore, the planning for the German sub-pilot is still in an early stage.

RBB is currently defining its requirements and researching online application/websites that would be most suitable for conducting the sub-pilot. It needs to integrate/ support a simple video player. Survey questions can be provided by paper.

RBB is also researching which content to use for the operational phase.

#### 4.1.2. Timeframe

The German Clean Audio sub-pilot is currently planned to start in M27 (February 2016), with an accompanying user evaluation. The sub-pilot will run for approximately six weeks. This needs to be further defined, depending on the prerequisite lab test methodology, conditions and outcomes.

#### 4.1.3. Test users and evaluation approach

For the sub-pilot RBB is planning to recruit at least 30 testers evenly spread between three levels of hearing impaired: slightly hearing impaired, moderate hearing and severe hearing impaired.

The panels would be as representative as possible and would be recruited with the cooperation of with the associations of deaf & hard of hearing of Berlin & Brandenburg. RBB would also contact participants to previous tests.

The testers would be using a PC and be provided with an URL to a website or a way to access the possible HbbTV application. They would be asked for some personal details such as hearing ability that are relevant for the evaluation of the results. Once the users have provided these details they will start the tests. This would involve playing a short video clip with Clean Audio. The user would then be required to complete a number of questions based on the clip they just heard. RBB intends to evaluate the users' satisfaction with the provided results of Clean Audio processing.

#### ***4.1.4. Challenges, problems and potential solutions***

In case the lab tests in autumn 2015 do not provide conclusive results for the use of stereo audio material, the sub-pilot will be strongly affected. In that case the large-scale test will need to be reconsidered, as most likely only limited 5.1 audio material can be made available. It would then be cooperatively decided, how the sub-pilot (possibly in another format) will be realised.

Other foreseen problems are more of an organisational nature. The aim of RBB is to recruit a demographically representative sample of testers. However as the testers may need to have an HbbTV device to take part in the sub-pilot, this may prove difficult. If this is the case RBB will look into the possibility of providing testers with equipment and adjusting the sample group demographics.

A further potential risk is that the testers do not understand the application or the methodology. For both we are being advised by local user associations thus minimizing this risk.

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

#### ***4.2.1. Status of work***

As explained in [2], TVC closely collaborates with UAB to perform tests with Clean Audio content with hard of hearing users in Catalonia, in controlled lab conditions. The CA content will be generated by IRT from selected TVC audio material by means of the CA generator. The results of the lab tests will be studied by TVC to evaluate the suitability of eventually developing a CA service, using IRT's CA generation component, thus implying taking into account the licensing of that component. In the future, content with the CA accessibility service could be published using the TVC's existing HbbTV VoD application "TV3 A la carta", depending on the results obtained from the CA sub-pilot with the targeted users.

The process of selecting suitable test content by TVC is still ongoing. The audio quality (in terms of dialogue intelligibility) of such content is critical: if the quality is too high, the improvement of the intelligibility after CA processing may not be noticeable during the evaluations. The CA content must therefore be "critical" with respect to intelligibility e.g. having dialogs that are difficult to listen compared to the background sound. UAB has started to conduct a survey amongst hard of hearing users to determine the types of contents that may be most suitable. A similar survey has already been performed amongst German hard of hearing users by IRT (see [8]) – the UAB/TVC survey builds on this experience and focusses on users in Catalonia.

The TVC workflow adaptation to work with assets with two audios has been finalised (see chapter 3.2). This allows the delivery of the original audio and the CA to end users, who have an HbbTV device.

#### ***4.2.2. Timeframe***

The timing of this sub-pilot strongly depends on the CA generator implementation, which is still ongoing, and another round of lab tests by IRT in autumn 2015. Also the selection of suitable content by UAB, based on the ongoing survey, has to be completed and test content created. Taking this into account, TVC foresees the following timeline:

- June / July 2015: survey about most convenient content for use in CA user tests.
- August / September 2015: gathering of suitable materials according to the survey results.
- October / November 2015: user tests conducted by UAB with the CA tracks generated by IRT.

- November / December 2015: evaluation of user tests and definitive decision about the performance of the CA sub-pilot. The following points depend on this decision.
- If finally it is decided to go ahead with the sub-pilot:
  - November 2015 / January 2016: generation of definitive CA contents to be used in the sub-pilot, preparation to be published and delivered as MPEG-DASH streams, and final adjustments to the HbbTV client reception application (for support of CA as an additional user option) that will be used by the users to test the accessibility services (CA, AD and other languages).
  - December 2015 / January 2016: Start of the CA sub-pilot in Catalonia.

### ***4.2.3. Test users and evaluation approach***

For the CA testing it is planned that potentially any user in Catalonia can take part in the sub-pilot, as long as he has an HbbTV 1.5 device which can run the desired functionalities of content playback and audio options selection of TVC's client application in a correct way. The evaluation will be done measuring and analysing the audience activity, which is planned to be done with in-app JavaScript event gathering using Adobe Omniture (as has been described in D4.1 [2]). This event gathering will measure the amount of requests of the application pages and "clicks" for each specific content and audio selection. On the event gathering TVC plans to include: application's page change (how much 'content lists' or player pages are requested), content selection (amount of views received by each content offered) and time of playback (how long users are watching the content), and audio selection (for a specific content item, how many times a concrete audio option – Catalan / CA / AD / other language - has been selected). Other means of feedback are not planned at the moment.

### ***4.2.4. Challenges, problems and potential solutions***

The selection of suitable content for CA evaluation is one of the biggest challenges. If the quality of the audio mix with respect to speech intelligibility is "too high", CA enhancements may go unnoticeable.

Also the "human factor" in the responsible production department – making production people understand the need of the additional effort to produce the CA contents and including it on the working chain – is challenging. The Clean Audio version may be different from the original audio track of a film or TV series, losing e.g. environment sound during dialogues, and some directors may find it an inadmissible loss in respect of the final result they wanted for their production. Convincing them of the changes on the audio content to get the CA, is a requirement.

If possible, TVC will combine all audio versions into one service offering (e.g. original language, CA, AD and a dubbed version). The extent to which this is possible depends on the availability of audio content. The rights issues involved also pose a challenge, as probably the most suitable content, being either a TV series or a movie, will not be owned by TVC. Thus, finding these contents may not be as easy as it could seem. At the time of writing, as contents for evaluating the accessibility services TVC has a Catalan TV series in property for which AD can be generated and a foreign TV series in English which can be offered in original as well as the dubbed Catalan version.

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

### 4.3.1. Status of work

TVC already produces Audio Description for broadcast of specific informative programs, by means of a Text-To-Speech solution. As described in section 2.2, an enhanced AD generation component to fit the HBB4ALL needs and requirements is currently under development. The workflow changes to carry the AD among the original audio, without description, in the same media asset content file has already been realised; the integration of the AD generation component in the TVC workflow is still pending.

TVC has been lab testing content with suitable Audio Description, generated in TVCs lab. Tests have been done with some HbbTV devices running a demo application, with satisfactory results. The final content to be used in the AD sub-pilot with the users has to be selected.

### 4.3.2. Timeframe

The AD asset generation and the beta testing of AD publishing, metadata generation and content playback has been fulfilled, without yet being fully integrated in the TVC workflow. All of them according to the MPEG-DASH standard, trying both VoD and live streaming of the contents and also delivering multiple audio assets, including AD in addition to the original audio. These tests have been performed in the TVC labs with HbbTV devices and developing an application for the playback of the content and the AD option selection by the user. All of them are working properly. The VoD streams have been tested as first beta since January 2015, and the live streams since March 2015.

The AD sub-pilot still has to be concretised in function of the improvements in the right publishing, delivery and client reception of the multiple audio MPEG-DASH contents. Its timing also depends of the final development and integration of the automatic AD generation component. TVC currently foresees following timeline for the AD Catalan sub-pilot:

- End September 2015:
  - Finalise AD generation component.
  - Generation of AD tracks of the final contents selected to be used in the sub-pilot, preparation of its publishing and delivery as MPEG-DASH streams, and final adjustments to the HbbTV client application.
- October 2015: start of the AD sub-pilot in Catalonia.

### 4.3.3. Test users and evaluation approach

For the AD testing it is planned that potentially any user in Catalonia can take part in the sub-pilot, as long as he has an HbbTV 1.5 device which can run the desired functionalities of content playback and audio options selection of TVC's client application in a correct way. The evaluation will be done measuring and analysing the audience activity, which is planned to be done with in-app JavaScript event gathering using Adobe Omniture (as has been described in D4.1 [2]). This event gathering will measure the amount of requests of the application pages and "clicks" for each specific content and audio selection. On this event gathering TVC plans to include: application's page change (how much 'content lists' or player pages are requested), content selection (amount of views received by each content offered) and time of playback (how long users are watching the content), and audio selection (for a specific content item, how many times a concrete audio

option – Catalan / CA / AD / other language - has been selected). Other means of feedback are not planned at the moment.

#### ***4.3.4. Challenges, problems and potential solutions***

The most important challenge of the performing of the AD sub-pilot is the availability of correctly working HbbTV 1.5 devices among the test users. Tests developed in the TVC labs showed that only a little amount of TV brands and models work as expected when running the functionalities of the HbbTV client application, though supposedly all devices under test according to their specification should support the HbbTV 1.5 standard features. Because of that, the potential amount of test users gets reduced to the ones who own device models that actually can properly run TVCs test application. This may strongly affect the actual user group (size, representativeness). Potential ways of managing this risk are currently under discussion.

As mentioned in section 4.2.4, TVC ideally would combine all audio versions into one service offering (e.g. original language, CA, AD and a dubbed version). The extent to which this is possible depends on the availability of audio content, and also rights issues play an important role, as suitable content may not be owned by TVC.

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

#### ***4.4.1. Status of work***

The TVC's workflow change to support multiple audio contents has been realised. The provisioning of additional language tracks was the original trigger for this, but as having files with audio contents in different languages is only an instance of offering different audio options to the users, it also has been used for the other Catalan sub-pilots. The other language generation, or more specifically the Catalan dubbed audio generation for the original language version, has already been done and is working, adapting the output audio formats to the required by the end devices.

TVC has lots of contents in different languages, commonly the original version in English and the dubbed audio in Catalan. This content has been used in technical lab tests, with HbbTV devices available at TVC, using publishing and delivery MPEG-DASH contents with two audio components, and playback and selection of a specific audio among these two components by TVCs HbbTV demo application. As in the case of the AD, these have all worked properly on a selected set of HbbTV devices.

#### ***4.4.2. Timeframe***

The foreign language asset generation and its publishing, metadata generation and content playback testing have been realised. The same features, requirements and conditions are relevant as explained for the AD sub-pilot; the same application is being used in both cases and satisfactory results in technical lab tests have been obtained. The VoD streams have been tested as first beta since January 2015, and the live streams since March 2015.

The other languages sub-pilot still has to be concretised in function of the improvements in the right publishing, delivery and client reception of the multiple audio MPEG-DASH contents. TVC currently foresees the following timeline for the other languages Catalan sub-pilot:

- October 2015: selection of the definitive contents to use in the sub-pilot, generation of the audio tracks dubbed in Catalan and preparation of these contents to publish and deliver them as MPEG-DASH streams. Final adjustments to the HbbTV client application.
- November 2015: start of the “Other languages and language learning” sub-pilot in Catalonia.

#### ***4.4.3. Test users and evaluation approach***

As described in section 4.3.3, the intended test users potentially are all users with HbbTV devices in the region of Catalonia, taking into account the market penetration of the HbbTV 1.5 capable devices. Unlike AD, foreign language is targeted at a broader audience and is expected to have more usage. The number of users that can benefit from this other languages service is a lot larger, as the target group does not need to have any hearing or viewing impairment. The evaluation will be done measuring and analysing the audience activity. This audience activity measurement will be done with in-app JavaScript event gathering using Adobe Omniture (as was already described in [2]). This event gathering will measure the amount of requests of the application pages and “clicks” for each specific content and audio selection. On this event gathering TVC plans to include: application’s page change (how much ‘content lists’ or player pages are requested), content selection (amount of views received by each content offered) and time of playback (how long users are watching the content), and audio selection (for a specific content item, how many times a concrete audio option – Catalan / CA / AD / other language - has been selected). Other means of feedback are not planned at the moment.

#### ***4.4.4. Challenges, problems and potential solutions***

Similarly as for the AD case, the availability of correctly working HbbTV 1.5 TVs among the test users is a big challenge of this sub-pilot. Only those users that own HbbTV devices (brands and model types) that have been successfully tested in TVCs labs (specifically with respect to support for the selection and playback of audio content using MPEG-DASH), could be the final test users of the sub-pilot. This may strongly reduce the potential user group (size, representativeness). Potential ways of managing this risk are currently under discussion.

As mentioned in section 4.2.4, TVC ideally would combine all audio versions into one service offering (e.g. original language, CA, AD and a dubbed version). The extent to which this is possible depends on the availability of audio content, and also rights issues play an important role, as suitable content may not be owned by TVC.

### **4.5. Overview of common challenges**

This section summarises the challenges with respect to the Pilot-B activities. Some relate to the realisation of the sub-pilots in the operational phase (Task 4.3 in the HBB4ALL project); some relate to the actual usage of the developed solutions in real life systems when assuming adoption of the new accessibility services by service providers such as broadcasters.

Challenge	Type	Sub-pilots <sup>4</sup>
It should be possible to integrate the new components without changing current audio workflows.	Technical / organisational	ALL
The potential integration in existing workflows should be as easy as possible.	Technical	ALL
Satisfy production personnel and management (quality issues)	Technical (and organisational with respect to the “human factor”)	ALL
Maturity of solution: Clean Audio generator (to be finalised), MPEG-DASH (relatively stable)	Technical	ALL
Penetration of HbbTV 1.5 in the market is relatively low. Testing of HbbTV based solution is only possible in a limited way. This also influences the potential user group (size, representativeness).	Technical / organisational	CAT CA/AD/LL
Evaluation methodology used to verify improved speech intelligibility. (Big challenge specifically for the evaluations in lab tests. Clean Audio is a novel solution which requires a novel evaluation method, which should also be easy to understand for all users).	Organisational	CAT CA, BB CA
Finding suitable content for CA evaluation in lab tests. If quality “too high” with respect to speech intelligibility, there may be no measurable effect.	Technical / organisational	CAT CA, BB CA
Getting rights for content	Organisational	CAT CA/AD/LL

**Table 1.** Overview of common challenges

<sup>4</sup> CAT=Catalan sub-pilot; BB=Berlin-Brandenburg sub-pilot; CA=Clean Audio; AD=Audio Description; LL=Other languages and language learning.

## 5. Complementary user tests

To support the requirements finding process and the fine-tuning of services and service components, technical as well as small scale “preliminary” user tests have been carried out. Within HBB4ALL user tests had 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 service requirements.

For Pilot-B a variety of preliminary user tests has been performed so far, some of which already have been documented in deliverable D4.1 ([2], section 7.1). An update is given in this chapter, with a description of further user tests until M20 in the project (see section 5.1). After a summary of the intermediate results thereof, section 5.3 provides an overview of further user tests planned to be carried out during the operational phase.

### 5.1. Preliminary user tests (until M20)

#### 5.1.1. Clean Audio

To evaluate the potential benefit of the Clean Audio generator for stereo signals, listening tests with both hearing impaired users with and without hearing aid were conducted. The results for 5.1 content already showed that a significant improvement in terms of speech intelligibility can be reached with the developed CA generator (see Deliverable D4.1, chapter 7.1.1 [2]). Hence, there is no need to conduct additional 5.1 tests.

Both tests in Munich (users having a light to middle hearing impairment without hearing aid) and Berlin (users having a severe hearing impairment with hearing aid) were carried out in March 2015 as per the recommendations of ITU-R BS.1284 [5].

As test setup a TV-set and stereo speakers was used. For both tests five sequences with a duration of approximately 10 seconds were selected from different TV shows. All of these sequences were tested in several different ways of CA processing and there was a hidden reference included as well. All in all there were 28 different sequences to evaluate. The order of the sequences was randomized and the same for all participants. Each sequence was played to the participants in the same way. First the original, unprocessed signal as reference, then the modified version and then both were repeated one more time. Therefore every block which the participants had to evaluate looked like this: reference - modified - reference – modified. After every block the participants had time to make their evaluation: they were asked to rate the speech intelligibility as well as the listening experience of the modified version compared to the reference signal on a scale from -3 (much worse) to +3 (much better); 0 indicated there was no audible difference between the two signals.

Although a very intensive familiarization was conducted with the test subjects before starting the test sequences, the test results showed that the participants had great difficulties to compare two versions (reference and modified) and to rate the difference between them rather than the intelligibility itself. Hence, most of the subjects were not able to rate hidden reference in an appropriate manner and had to be rejected (a score of 0 for the hidden reference would be ideal but even for small deviations (+/- 1) only a few subjects passed the post screening).

Moreover, the participants reported that some of the reference clips were not critical and in their original version already had a good speech intelligibility. Hence for such clips no improvement can be reached by the CA generator.

This situation lead to the conclusion that additional tests have to be conducted with a more appropriate evaluation methodology in order to have a good starting position with respect to the creation of audio material for the CA sub-pilots. These additional tests are planned for autumn 2015 (see section 5.3.1.1).

### 5.1.1.1. Lab test, Munich

A lab test with 18 subjects was conducted at IRT premises in March 2015. All of the test participants had a light to middle impairment of hearing and did not wear a hearing aid.

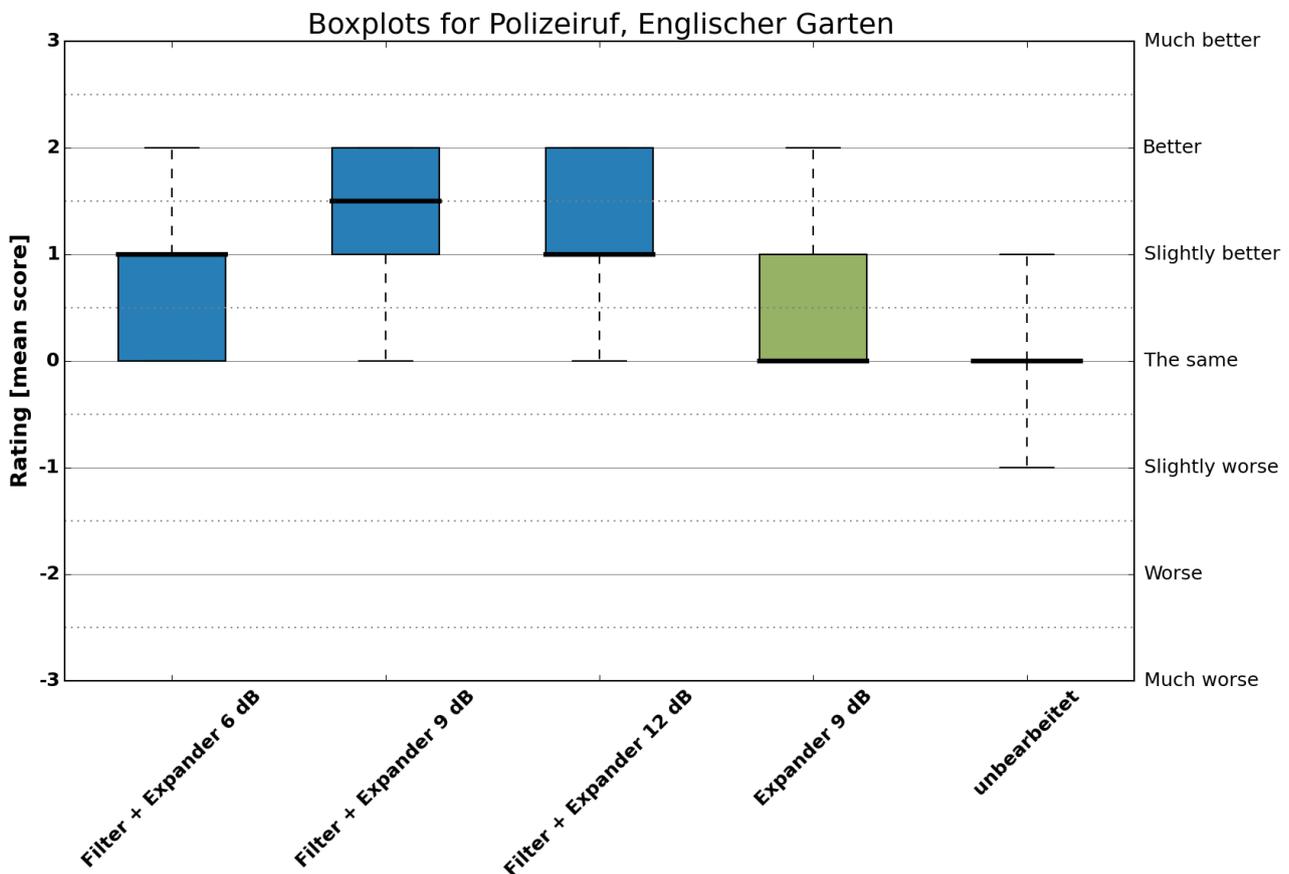


Figure 7. Results for one specific test clip of the listening test with hearing impaired people that do not wear a hearing aid

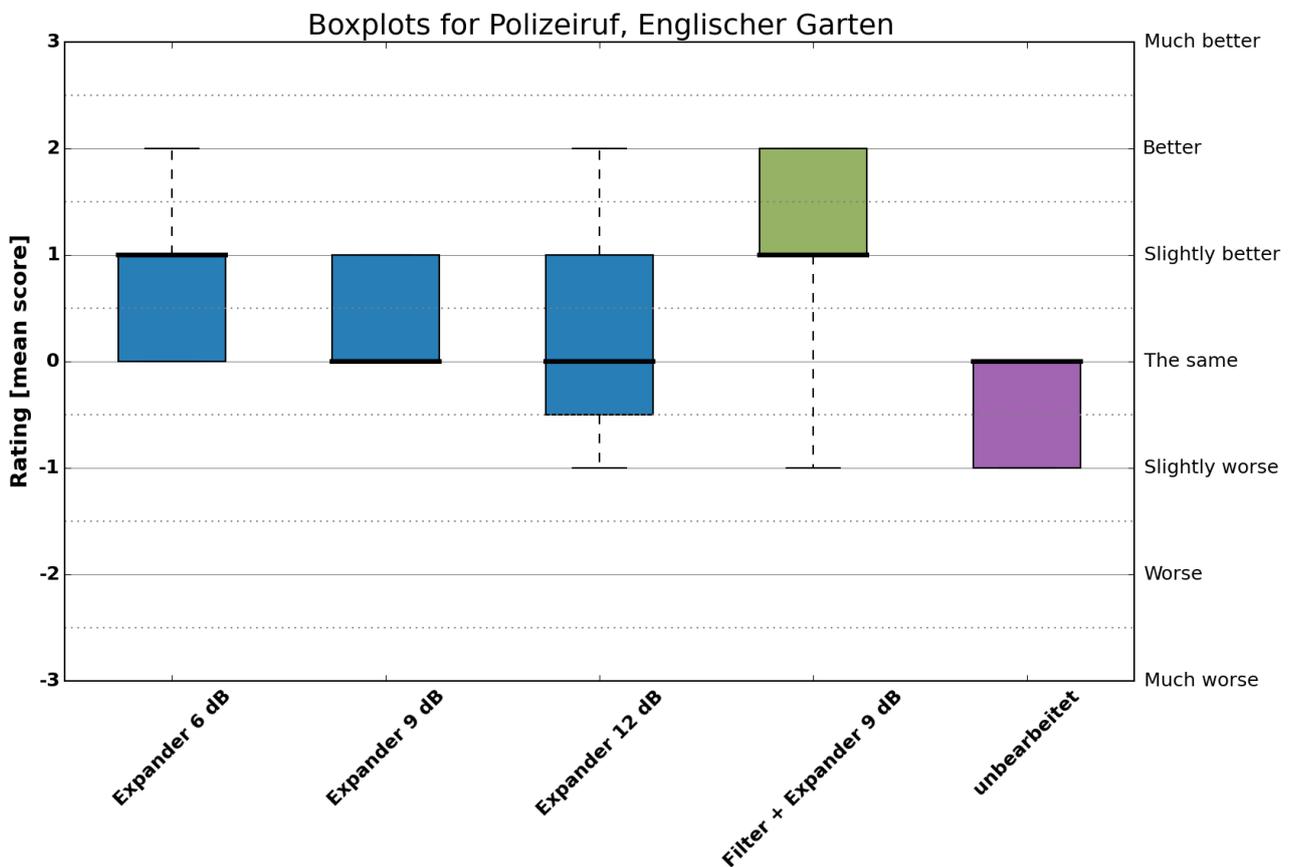
Although the results of this listening test are not completely reliable (12 out of 18 participants didn't pass the post screening), trends can be seen from specific test items. The tested settings and combinations of the CA generator reached an improvement for some test items but not for all of them. Specifically, processed versions of one test item (which has been reported as the only one with bad speech intelligibility) reached a significant improvement in terms of speech intelligibility. Figure 7 illustrates the results for this specific clip as boxplots. The 'unbearbeitet' (English: "unprocessed") result (most right) is for the hidden reference ratings.

However, for other test clips no significant improvement could be reached.

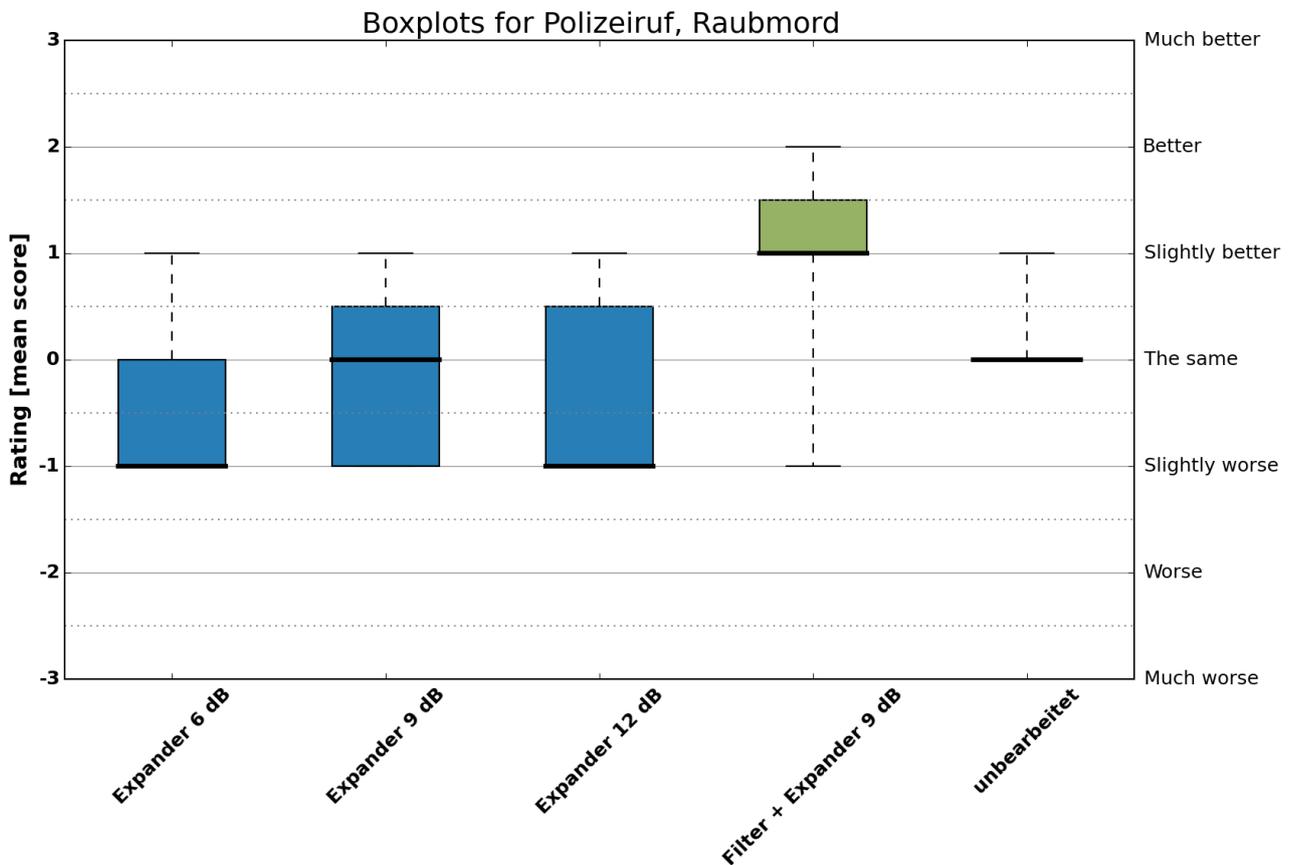
### 5.1.1.2. Focus Group, Berlin

A further test was conducted at RBB in Berlin with 28 participants, also in March 2015. Especially the subjects from this test had great difficulties to identify the hidden reference and therefore 17 of them had to be excluded from the final results after post screening. For the 11 remaining participants, two tested clips showed significant improvements. Interestingly, one of them was the same as for the test at IRT. **Figure 8** and **Figure 9** illustrate the results for those clips. The version “Filter + Expander 9dB” for both clips reached the best result.

Other tested clips didn’t reach a significant improvement in terms of speech intelligibility. One reason might be that the content was not ‘critical’ enough with respect to speech intelligibility.



**Figure 8.** Results for one specific test clip of the listening test with hearing impaired people wearing a hearing aid



**Figure 9.** Results for one specific test clip of the listening test with hearing impaired people wearing a hearing aid

### 5.1.2. Audio Description

Until M11 of the project, the following tests had been completed by UAB:

- Second Screen Application for Audio Description
- Other languages and language learning:
  - Audio Description as a tool for language learning
  - Audio Description as a didactic tool

For specific information regarding the above mentioned tests as well as main results obtained, please consult Deliverable D4.1 ([2], sections 7.1.2 and 7.1.3).

UAB has planned further tests in the context of Audio Description. They are either currently in progress (see sections 5.1.2.1, 5.1.2.2 and 5.1.2.3 hereof, and then further section 5.3.2 hereof) or in planning and preparation stage (see section 5.3.2 hereof).

### ***5.1.2.1. Genre-specific Audio Description***

User tests already have been conducted in Poland. After detailed analysis of the results, it turned out that the sample was too small to give significant results, and therefore additional tests were scheduled for August/September 2015.

Aim: check preferences and levels of presence of Audio Description users for neutral vs. alternative version of Audio Description.

Methodology: the vehicle for the experiments was the feature film ‘The Mighty Angel’ (2014, dir. by W. Smarzowski). For the purpose of the study, two Audio Description scripts were prepared: one with emotive references taken from the screenplay and the other being more detached, which followed the existing tradition of “what you see is what you say” rule.

Participants:

- stage 1: 15 blind and partially-sighted participants involved in the tests (they were recruited through associations working with the visually impaired people in Poland);
- stage 2: call for participants has already been made to Polish user associations working with the visually impaired people; participants are being recruited.

Hypotheses:

- 1: Audio Description users will show higher levels of presence for alternative version of Audio Description;
- 2: Audio Description users will show a preference for alternative version of Audio Description.

Main conclusions (after the first stage of tests):

- users preferred an alternative version of Audio Description carrying more emotions than traditional (objective) Audio Description;
- levels of presence are to be analyzed further.

### ***5.1.2.2. Vocal delivery of Audio Description by Film Genre***

User tests already have been conducted in Poland. After detailed analysis of the results, it turned out that the sample was too small to give significant results, and therefore additional tests were scheduled for August/September 2015.

Aim: examine the immersion in filmic content (emotive vs. informative) with Audio Description delivered with human and synthetic voices, and verify user preferences as for the choice of voice for Audio Description delivery by genre.

Methodology: the vehicles for the experiments were fragments from the feature film ‘Charlie and the Chocolate Factory’ (2005, dir. by T. Burton) and the documentary film ‘Encounters at the End of the World’ (2007, dir. by W. Herzog). For the purpose of the study, Audio Description scripts to the above mentioned film fragments were prepared and recorded with human and synthetic voices.

Participants:

- stage 1: 15 blind and partially-sighted participants involved in the tests (they were recruited through associations working with the visually impaired people in Poland);
- stage 2: call for participants has already been made to Polish user associations working with the visually impaired people; participants are being recruited.

Hypotheses:

- 1: Audio Description users will show higher levels of presence for human Audio Description for a feature film, but little or no difference in the levels of presence between human and text-to-speech Audio Description for a documentary;
- 2: Audio Description users will show a preference for human Audio Description for a feature film, but no preference between human and text-to-speech Audio Description for a documentary.

Main conclusions (after the first stage of tests):

- Audio Description users showed a preference for human Audio Description for a feature film, but no preference between human and text-to-speech Audio Description for documentary;
- levels of presence are to be analyzed further.

### ***5.1.2.3. Cultural Allusions & Intertextuality***

User tests already have been conducted in Italy and the UK; additional tests are in planning to be conducted in Spain; please see section 5.3.2.1.

Aim: compare English and non-English ADs for the same film, with English and non-English audiences; check engagement and comprehension

Methodology: the vehicle used for the experiments was the feature film ‘The King’s Speech’ (2010, dir. by T. Hooper). It is a British historical drama, set in the glamorous world of British royalty, and thus carrying a characteristic British “flavour”. It was available in the original language of English and in its dubbed versions in Italian and Spanish. Similarly, audio described versions existed in all three languages.

Participants:

- two British audiences: one a group of normally sighted students at the University of Loughborough and the other a group of visually-impaired persons (VIPs) at the Royal National Institute for the Blind (RNIB) in London

- two Italian audiences: normally sighted audience of students at Trieste University and a group of VIPs at the Unione dei Ciechi (Blind Union) in Udine.

Hypothesis: access to cultural allusion and intertextuality will increase engagement and comprehension (the English audience expected to report greater levels of both than the non-English audiences)

Conclusions (after the tests in the UK and Italy):

- In terms of enjoyment all the RNIB participants claimed that they very much enjoyed the film and most said they definitely enjoyed the particularly British “flavour”. One hundred per cent of the Loughborough students said they enjoyed the film and the British “flavour”.
- From the Italian groups the Unione Ciechi participants also agreed that they enjoyed the film but were more reticent about its British “flavour”, though nobody criticised this aspect. The Trieste students responded in a broadly similar way apart for one participant who did not appreciate the strong British “flavour”. So at least as regards enjoyment it cannot actually be said that the foreign audience liked the film less than the home viewers.
- Turning to comprehension, a number of differences began to emerge. Among both the VIP and the non-VIP groups cultural allusions were seen to be picked up more easily among the English especially, and understandably, with regard to language markers of social and geographical extraction. This was particularly the case with the older participants in the RNIB group. Interestingly, no great differences in the understanding of the cultural elements emerged between the two groups of students, though it must be pointed out that the Trieste contingent were also language students and this may definitely have skewed the results. Even the quiz results, while showing higher scores among the English students, did not point to a huge divide whereas in the case of the VIPs, this was much more evident. In fact, with these groups it was the quiz in particular that proved to indicate that, although it appeared to have no effect on enjoyment, there was a noticeable lack of culturally-based knowledge on the part of the foreign audience.
- So, while the original hypothesis that “access to cultural allusion and intertextuality increases engagement and comprehension and that, as a consequence, an English audience watching an English film would report greater levels of both than non-English audiences” appears to have been confirmed as regards the VIP groups involved, it did not seem to impinge on the audience’s enjoyment of the film. And, serendipitously, the tests showed that this was not necessarily the case, or not to the same extent, when considering the non-VIP audiences of university students.

## 5.2. Summarised results and guidelines / recommendations

This section provides an overview of all user test results obtained within Pilot-B so far; some were already covered in detail in D4.1 ([2]), the recommendations of those tests are also included here. **Table 2** gives an overview of the recommendations as obtained from CA lab tests, regarding the use of Clean Audio specifically in user tests and the sub-pilots in Berlin-Brandenburg and Catalonia. **Table 3** summarises the recommendations resulting from the AD user testing, describing possibilities to improve AD services in general, from an end-user point of view, also beyond Pilot-B.

Issue	Clean Audio recommendation
Subjective evaluation methodology	<p>Subjective evaluation as per ITU-R BS.1284 does not guarantee relevant results.</p> <p>Especially for elderly and hearing impaired persons, a more suitable methodology should be applied in lab test evaluations to achieve rather significant and relevant results. In the new method audio test sequences should only be played once, to prevent an increase in speech intelligibility by repeated inputs ('memory effect').</p>
Input content format for the CA generator	<p>The conducted tests clearly showed that an improvement in terms of speech intelligibility can be reached with 5.1 input format. For stereo content, it depends on several factors. Hence, if 5.1 content is available, it should be used for the CA generator.</p>
Test content for subjective evaluation	<p>The source material for subjective evaluations should be very carefully selected w.r.t the expected improvement in speech intelligibility. Only critical test items (with bad dialogue intelligibility) can demonstrate an improvement by the CA generator.</p>
Users wearing a hearing aid or not	<p>The current results show that it is sensible to distinguish (in lab tests as well as in a potential service provisioning) between people wearing a hearing aid and those who do not. Due to the audio processing in hearing aids, the optimal CA parameter settings may be different between these user groups.</p>
In case only one additional CA stream is provided to the end-user (e.g. as part of a broadcaster's service)	<p>The current results show a trend: the highest improvement in terms of speech intelligibility can be reached with the "Filter+Expander 9dB" combination for both user groups, with and without hearing aids. This combination should be considered for a single stream service.</p>
CA as server side solution	<p>Ideally, audio mixing under end-user control would be done in the receiver. Current receiver technology</p>

	<p>and audio production workflows do not support such a solution.</p> <p>Therefore, as a compromise between full personalization and technical feasibility, it is recommended to offer several pre-produced CA versions with different mixing levels to the user.</p>
--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Table 2.** Overview of recommendations with respect to Clean Audio

Case Type	Recommendation with respect to Audio Description
Films/programs where male voices are in the majority	Deliver AD with a female voice
Films/programs where female voices are in the majority	Deliver AD with a male voice
Films/program with a mix of voices (male and female)	Deliver AD with a male or female voice, possibly different from those in the original film soundtrack
Feature films	Deliver AD with a human voice
Documentary films	Deliver AD with a human or a synthetic voice
Films/programs for bilingual audiences	Deliver AD in the same language as the language of the film (e.g. film in Catalan with AD in Catalan; film in Spanish with AD in Spanish)
Foreign films with subtitles	<p>In case subtitles are recorded as audio subtitles, use a voice different than the one used for AD, e.g.</p> <ul style="list-style-type: none"> <li>• if AD is already delivered with a female voice, then audio subtitles should be delivered with a male voice and vice versa, or</li> <li>• if AD is already delivered with a female voice, then audio subtitles may be delivered with a female voice, but different than the</li> </ul>

	one used for AD
Films with dense dialogues	Complement AD by AI
AD style	Deliver AD with a style that matches the style of the film; not necessarily objective and flat
AD voice	Deliver AD with a voice that matches the style of the film. (The voice must be adapted to the emotional content of the scene and the pace of action but must nevertheless remain fairly neutral. If an actor with too strong a personality is used, he will be competing with the actor in the film.)
AI style	Deliver AI with a style that matches the style of the film; may contain additional information on the plot, actors, film background
AD track for mobile applications	<p>To be played in mono, not stereo (AD track should not be put in stereo mode as the blind and partially sighted viewers use only one earplug for listening to it; they also have to follow the original soundtrack of the film played in the movie theatre).</p> <p>Proper sound mix required (volume of AD track should be adjusted to the volume of the film soundtrack).</p>

**Table 3.** Overview of recommendations with respect to Audio Description

## 5.3. User tests outlook

### 5.3.1. Clean Audio

#### 5.3.1.1. Lab tests (Germany)

As described in section 5.1.1, the chosen methodology and test items in the CA lab tests carried out so far, did not allow to conclusively derive generic and stable knowledge. As it is crucial to have reliable test results for the planned sub-pilots, an additional lab test with stereo content with a new testing methodology is required.

The first and important step will be to identify a large set of critical test items. This will be done with a small group of hearing impaired persons which will rate short clips regarding their speech intelligibility. These clips will then be the test pool for the additional listening tests at IRT and/or RBB.

Moreover, the identified weak points of the last testing methodology will be taken into account to design a more suitable test. Especially the repetition of a test clip should be avoided as the memory effect obviously influences the ratings for a speech intelligibility test. IRT and UAB will cooperatively design a new testing methodology.

This additional test is planned to be conducted until end of October 2015. This will leave enough time for the evaluation and preparations for the sub-pilots, and still allow the sub-pilots to be carried out within the foreseen timeframe of T4.3. The number of participants should be comparable to the last test in March 2015. Both groups hearing impaired with and without a hearing aid will need to be tested again separately.

### 5.3.1.2. *Lab tests (Spain)*

TVC, in cooperation with UAB, will select and cut some relevant TV show and film clips to test the advantages CA can provide to its audience. The relevance of this selection will be drawn by a questionnaire designed for the audience to ascertain which types of TV programmes are less intelligible, thus more prompt to benefit from the CA service. This builds on a similar questionnaire which was part of HBB4ALL activities carried out by IRT and RBB in Germany<sup>5</sup>, now focusing on users in the region of Catalonia.

These clips will be then processed with the CA generator and shown to a sample of four different target audiences in several lab tests in order to draw conclusions about:

- 1) how hard of hearing and the “general audience” consider the CA implementation with respect to speech intelligibility (this section);
- 2) how CA affects the likeability and comprehension of Audio Description for visually impaired people (for a detailed test description see section 5.3.2.3);
- 3) and whether CA makes second language acquisition easier or not (for a detailed test description see section 5.3.3.1).

Regarding the CA user tests with hard of hearing, as described in section 5.1.1, the methodology needs to be adapted. At the time of writing, IRT and UAB are putting together different ideas and approaches, in order to carry out evaluations using the same methodology both in Catalonia and Germany, to have a broader sample and more reliable results. The exact procedure for the methodology has not yet been finally decided. Generally speaking, the outline will be playing a number still unknown of clips, some of them with “normal” audio (‘reference’) and some of them processed with CA (using different parameter settings). The order in which we show them to each group will be randomized and it will not be disclosed which ones have been processed. As described in section 5.1.1, the method will definitely avoid repetition to prevent the memory effect to influence the ratings for a speech intelligibility.

After the viewing, UAB plans to pass on a questionnaire with different types of questions that help ascertain if CA improves in any way the participants’ experience. UAB will be able to prove or refute these hypotheses by making the sample audiences to fill out a questionnaire including recall, stated preference, and/or vocabulary questions on it. Recall questions are aimed for checking if the subjects remember both parts of the clip and their reactions to it; stated preference questions are designed to get further insights about

<sup>5</sup> [http://www.schwerhoerigen-netz.de/RATGEBER/RUNDFUNK-FERNSEHEN/PDF/erg1\\_umfrage.pdf](http://www.schwerhoerigen-netz.de/RATGEBER/RUNDFUNK-FERNSEHEN/PDF/erg1_umfrage.pdf)

the subjective perception of the viewers; and vocabulary questions will be useful to test the influence of CA in second language learning (see also section 5.3.3.1).

### **Clean Audio for hard of hearing**

**Aim:** To test whether CA-processed clips are more intelligible for hard of hearing viewers.

**Method:** The participants will be shown selected clips without subtitles and then answer some recall and stated preference questions. A questionnaire with recall questions the experiment will reveal whether viewers could understand the contents enough for them to remember them; while stated preference questions try to gather people's information about if they prefer a Clean Audio track or the usual sound for each specific clip.

**Participants:** 30 hard of hearing people, with and without hearing aids or cochlear implants.

**Hypothesis:** CA will facilitate intelligibility for the hard of hearing, by providing a more clear and understandable dialogue track without (or with less) accessory noises and sounds that may complicate its comprehension.

### **Clean Audio for the general audience**

**Aim:** To determine if people without visual or hearing impairment (who are not necessarily learning another language) will consider CA useful or, at least, inconsequential. It has to be taken into account that many families include both hearing and hard of hearing members, therefore there is a need to verify whether one member of the family watching a CA product will disturb those members who do not need it.

**Method:** Participants will fill out a questionnaire consisting on stated preference questions.

**Participants:** 30 participants without any hearing or visual impairment will be shown the clips in their own language.

**Hypothesis:** The experience of the hearing and sighted audience, if not improved by the increased intelligibility achieved by CA processing, will not be worsened either.

## ***5.3.2. Audio Description***

UAB has planned further tests in the context of Audio Description to be carried out in the project. For detailed information, please see the sections below.

### ***5.3.2.1. Cultural Allusions & Intertextuality***

Partial tests have already been conducted in Italy and the UK (for a detailed test description refer to section 5.1.2.3 hereof). The next part of these tests is to be carried out in Spain in July and August 2015. The test method and aims for the Spanish tests are the same as for the tests already carried out in the previously mentioned countries.

The tests are in the preparation stage at the time of writing:

- filmic materials have been prepared (Spanish version of “The King’s Speech” with Spanish Audio Description);

- questionnaires have been adapted for Spanish end users (they have been translated from Italian and English into Spanish);
- participants are being recruited (Spanish Association “DOCE” for persons with visual impairments has been contacted; call for volunteers has been published among Association members and in social media networks).

### 5.3.2.2. *Audio Introductions*

Tests are in the preparation stage at the time of writing (questionnaires are drafted, filmic materials are being processed and participants recruited).

The tests are to be conducted in the UK and are planned within the timeframe of September-December 2015.

Aim: examine the presence, interest and comprehension of the filmic content delivered with AD and Audio Introduction (AI) and with AD only; verify user preferences for human and text-to-speech voices.

Hypotheses:

- 1) AD users watching with AI expected to show higher levels of presence, interest and comprehension than watching without AI;
- 2) participants expected to show little or no preference for human over text-to-speech delivery of AI.

### 5.3.2.3. *Clean Audio and Audio Description*

Clean Audio is targeted for viewers with hearing impairments, but may as well serve as improvement for listening in noisy environments. Audio mix with fixed balance between dialogue and background is always a compromise and therefore finding the right mix can pose a challenge. Bearing this in mind, end user tests will be carried out as described in section 5.3.1.2. However, another application of CA – for visually impaired audiences – could also prove to be useful. The following test aims to verify that.

Aim: to test whether the combination of Audio Description and “cleaned” dialogues improves the understandability of the selected clips for the blind and visually impaired audiences.

Method: the Audio Description band of a product is already cleaned and does not need more processing in order to become more intelligible and accessible. Nevertheless, the intelligibility of the (original) dialogues is not often so, thus this test will show a clip with both Audio Description and CA-processed dialogues and confirm our hypothesis aided by some recall and stated preference questions. This is probably the most subtle and hard to determine of all CA applications since we will be handling an already cleaned AD track and checking if the combination of such with a CA dialogue track actually makes any difference for the user, therefore the testing of the applicability in this workflow might prove to be more complicated.

Participants: UAB will test 30 visually impaired people whose native language is Catalan. This lab test will take place in cooperation with some Catalan organizations for the blind and visually impaired.

Hypothesis: the clips combining “cleaned” dialogues and Audio Description are more easily understood if no unnecessary noises prevent the full attention of the audience.

### 5.3.2.4. *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, this project is restricted to sound quality in audio descriptions of pre-recorded content to be TV broadcasted. This restriction is due to two main reasons: one, TV is the most common channel of audiovisual content consumption since 97% of households had a TV and regard it as basic hardware, and second, 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 this project is that a “safe area” in terms of hearability can be defined to guarantee that the audio description narration is always heard correctly in any audiovisual product. An adequate sound post production process can guarantee that these “safe area” is always maintained, to the same extent as an inadequate sound mix could seriously threaten the quality of any audio description.

It has to be noted that there is no specific technical standard on sound mixing or volume with regard to audio description, and therefore no real quality control can be established over pre-recorded broadcasted TV content. This results in the fact that quality standards or even the feasibility of broadcasted audio description cannot be guaranteed for the visually impaired audiences. The British “ITC Guidance On Standards for Audio Description” (see: <http://www.ofcom.org.uk>) were the first guidelines to mention the volume and mixing processes used for the purposes of Audio Description and what they offer is deemed to be the “code of best practice” to be followed in the field.

UAB has planned a test aiming at establishing objective values on volume and sound mixing to be applied in broadcasting to guarantee a minimum quality for Audio Description sound levels. To this end the following experiment has been designed. Three clips, ca. 2 minutes each, from “Inglorious Basterds” (2009, dir. by Quentin Tarantino) will be tested in 3 scenarios, namely:

- scenario 1: Audio Description track normalized at -23 LUFS<sup>6</sup>; then inserted in the clip without modifications;
- scenario 2: Audio Description track normalized at -23 LUFS; then inserted and, where needed, the original OST lowered so that Audio Description can be heard properly;
- scenario 3: Audio Description track normalized at -23 LUFS; then inserted and adjusted to the volume fluctuations of the OST, which is left without modifications.

The clips have been chosen based on significant sound differences within each 2-minute scene. Comprehension and user preferences for different scenarios (audio mixing processes of the each clip) will be verified.

Expected outcome: users are expected to show preferences on each mixing process applied, which, in turn, is supposed to confirm the recommendations on volume and mixing processes used for the purposes of Audio Description provided for in the British guidelines (ITC Guidance On Standards for Audio Description).

The test is currently under preparation: Audio Description is being created for the purposes of the experiment; filmic clips are to be cut and mixed for different experimental scenarios; the questionnaires are being drawn up.

---

<sup>6</sup> LUFS – Loudness Unit referenced to Full Scale. The value of -23 LUFS is the current loudness target level for broadcasting in Europe under the EBU R128 (see: <https://tech.ebu.ch/docs/r/r128.pdf>)

### ***5.3.3. Other languages and language learning***

#### ***5.3.3.1. Impact of Clean Audio in language acquisition***

UAB plans to test CA for the second language learners as well, in order to determine the scope of this technique.

Aim: to check if new vocabulary items are more easily assimilated when the respective clips are more intelligible (i.e. processed by the CA generator).

Method: in order to validate the hypothesis that “cleaned” dialogues make it easier for language learners to isolate new items of vocabulary, UAB will design a recall and vocabulary questionnaire. If the hypothesis of CA helping with second language acquisition is proven, the lab tests could be extrapolated to the second language learning broadcasts and classrooms. Moreover, even when the programmes are not designed for second language acquisition, CA could make a difference for people watching the original versions of films and TV shows (original meaning: not their mother tongue).

Participants: 15 Catalan-speaking adults will be tested with clips in English, and 15 Spanish-speaking children will be tested with clips in Catalan.

Hypothesis: CA may improve language learner’s experience by reducing the effort they have to make in order to understand media in another language.

## 6. Conclusions

Work Package 4/Pilot-B has advanced well, and in general is performing in line with the project plan; the single exception is the work on Clean Audio, which is somewhat delayed, see below. Since the writing of D4.1, the remaining work for both tasks T4.1 and T4.2 has progressed well: all service components have been finalised or will be completed shortly; this is also true for the integration of workflows for use in the sub-pilots and the preparation of the various sub-pilots.

An additional lab test for Clean Audio with stereo content is required, as the chosen methodology and test items so far did not allow to conclusively derive generic and stable knowledge. This additional test is planned to be conducted until end of October 2015. This will leave enough time for the evaluation and preparations for the sub-pilots, and still allow the sub-pilots to be carried out within the foreseen timeframe of the Pilot-B operational phase (in T4.3). The main objectives for Pilot-B can thus still be achieved and no further specific measures, actions or contingency plans are considered necessary at this stage.

A large amount of preliminary user testing has been carried out until M20 of the project:

- With respect to Clean Audio, the test results have been used to adapt our further testing (e.g. evaluation methodology, test content to be used). Also they will be considered for the respective CA sub-pilots (e.g. input content, parameter settings.).
- With respect to Audio Description, a variety of results describes possibilities to improve AD services in general, from an end-user point of view, also beyond Pilot-B.

The partners in WP4/Pilot-B have cooperated strongly, both on organisational issues as well as on technical developments:

- CA user testing was carried out in sessions in Berlin and Munich cooperatively by IRT and RBB;
- Novel CA tests have been prepared (e.g. aiming for harmonized testing methodologies) by UAB with support of IRT;
- AD user testing was carried out by UAB, in close cooperation with TVC, also with respect to the preparation of further tests, selection of content, etc.;
- Technical developments for use of multiple audio delivery by MPEG-DASH as well as for selection of an audio asset in an HbbTV application have been coordinated and tested between TVC and IRT.

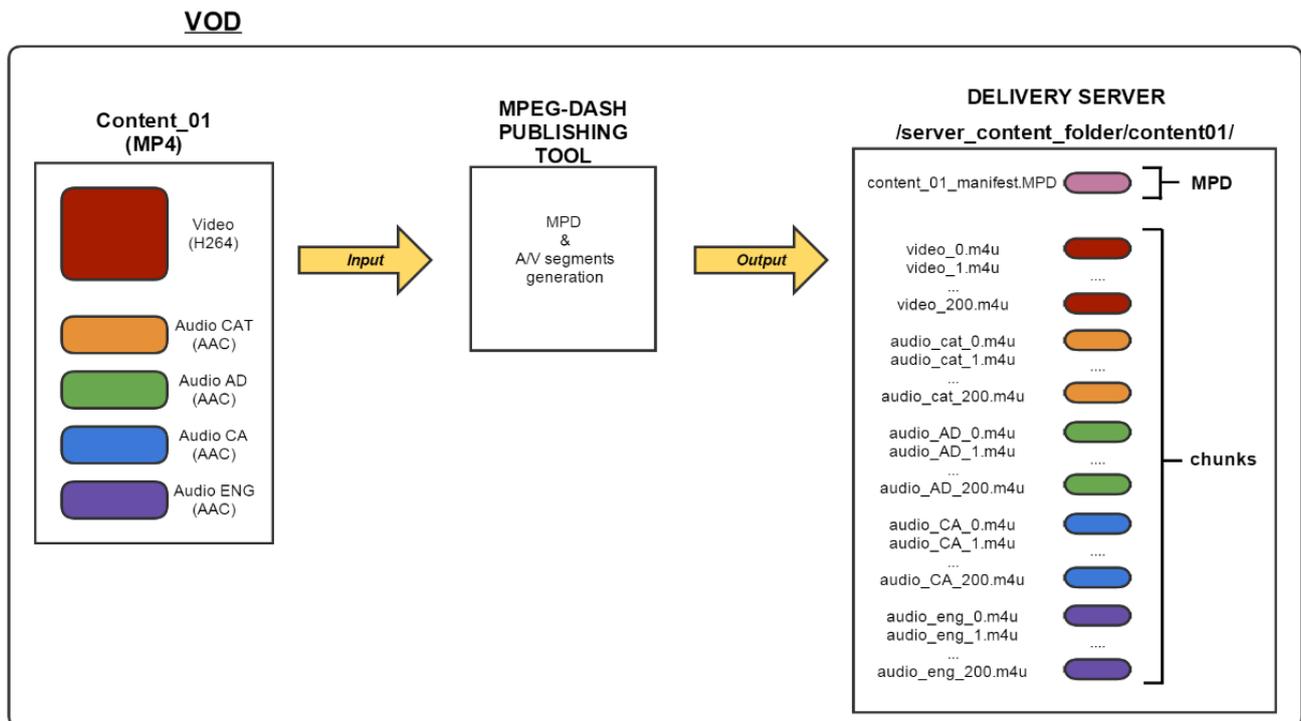
## 7. References

- [1] D2.3.2 – *Common Technical Components (II)*, HBB4ALL deliverable, 2015
- [2] D4.1 – *Pilot-B Progress Report*, HBB4ALL deliverable, December 2014
- [3] The “CenterCut” algorithm: <http://www.virtualdub.org/blog/pivot/entry.php?id=102>
- [4] ITU-R BS.775-3 – *Multichannel stereophonic sound system with and without accompanying picture*
- [5] ITU-R BS.1284 – *General methods for the subjective assessment of sound quality*
- [6] OIPF – Release 2 Specification – Volume 5 - *Declarative Application Environment – V2.3*, Open IPTV Forum specification, 2014-01-24
- [7] ISO 639-2 – *Codes for the representation of names of languages - Part 2: Alpha-3 code*, International Organization for Standardization standard, 1998
- [8] *Online-Umfrage zur Sprachverständlichkeit von TV-Produktionen für Hörgeschädigte* (results of an online survey on dialogue intelligibility of TV productions for the hard of hearing), IRT / RBB March 2015, [http://www.schwerhoerigen-netz.de/RATGEBER/RUNDFUNK-FERNSEHEN/PDF/erg1\\_umfrage.pdf](http://www.schwerhoerigen-netz.de/RATGEBER/RUNDFUNK-FERNSEHEN/PDF/erg1_umfrage.pdf) (all participants), [http://www.schwerhoerigen-netz.de/RATGEBER/RUNDFUNK-FERNSEHEN/PDF/erg2\\_umfrage.pdf](http://www.schwerhoerigen-netz.de/RATGEBER/RUNDFUNK-FERNSEHEN/PDF/erg2_umfrage.pdf) (participants wearing a hearing aid) (both documents in German only)

## 8. Annexes

### 8.1. Sub-pilots in Spain (Catalonia) – publishing and delivery for VoD

For the VoD case, the MPD and all the chunks for the total duration of the media asset content file are stored and are always available in the MPEG-DASH media delivery server directory. To generate them, TVC will most likely continue to use the GPAC's MP4box tool, which has been used up to now in HBB4ALL tests and demos. The separated video and audio components of the content are input to MP4box and the parameters for the segment duration are set, as well as the MPEG-DASH profile or segment template among all the attributes of the MPD that need to be determined. The generated MPD has every A/V component in a separated Adaptation Set, one for the video and one for each audio, to allow the end user to switch between the different audio streams on his HbbTV device. In **Figure 10** a picture of the VoD publishing/delivery workflow at TVC is provided.



**Figure 10.** Publishing/Delivery workflow for VoD streams

The duration chosen for the chunks in this case is 4 seconds. During tests TVC found that, if the video and audio chunks did not have the same duration, in some devices video and audio got out of synch when playing VoD MPEG-DASH streams of long duration. This problem can be avoided when generating video and audio chunks with exactly the same duration. To accomplish this, without changing the 1 second video GOP size, the audio sampling frequency was set to 32 KHz (32000(samples per second)/1024(frames per AAC sample) = 31'25 frames per second → 32 ms per audio frame x 125 = 4 seconds). Several combinations of audio sampling rate were analysed which could generate audio chunks at a multiple of the GOP size, and 4 seconds was the shortest possible segment duration; this way both audio and video segments can have the same duration accurately avoiding any synchronization problem.

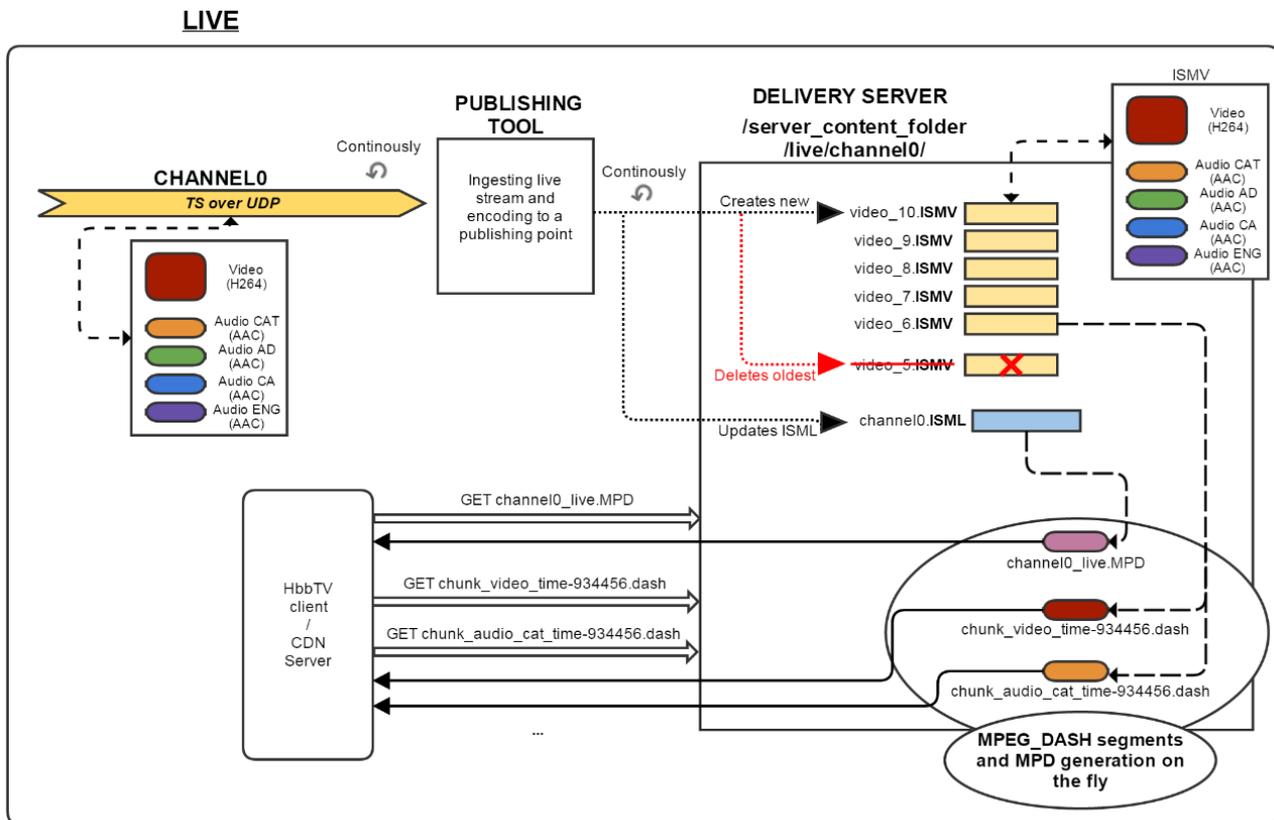
Here is an example of codec settings of the contents tested with satisfactory results in different HbbTV client devices:

- Video → Higher quality: 7Mbps, 1920\*1080 (16:9), 25fps, AVC (High@14.0), GOP size = 1s.
- Audios: → 128 kbps, 32 KHz, 2 channels, AAC (LC) (in 4 components: CAT (Catalan, dubbed), AD, CA, ENG (English, original language))
- MPD: number based.
- Chunks: 4s, 4\*GOP size.

For the hosting and delivery of the MPDs and chunks to the clients using HTTP URL's TVC uses the free and open-source Apache HTTP Server, on which the mime type application/dash+xml was configured for the MPD extension. With the public IP of the Server, and the listening port, the DASH contents placed in the specific folders can be pointed, requested and delivered to the users, through a CDN.

## 8.2.Sub-pilots in Spain (Catalonia) – publishing and delivery for live IP-streaming

For the live case TVC uses Unified Streaming Platform (USP). This product works with its own manifest and segment files, which for live streaming are ISML, called publishing point, for the description of the stream being delivered and ISMV for the chunks containing the audio and video contents. These files are stored in an MPEG-DASH media delivery server folder, and the content information of the ISML and the ISMV segments with the content itself are updated with the ffmpeg tool. With ffmpeg a Transport Stream (TS) is ingested containing the multiple audios and the video, transported via UDP, and from the TS the ISMV fragments are generated, see **Figure 11**. The ISML file is always stored in the server and updated over time, whilst the ISMV files are only stored and available for a certain amount of time: the content is generated live and segments corresponding to an instant of time already passed are deleted while new ones containing the current instants of time of the stream being ingested are created. The amount of total time of the ISMV fragments stored in the server is configurable.



**Figure 11.** Publishing/Delivery workflow for Live streams.

In the live scenario explained above, the MPD and the specific MPEG-DASH video and audio segments are never stored in the server folders. All of them are generated dynamically, on the fly, when a client requests them. The MPD is generated from the information available in the publishing point file, ISML, and the chunks are created from the content of the ISMV fragments. The delivered MPD has every A/V component in a separated Adaptation Set, one for the video and one for each audio, as desired to allow the clients to switch between audios.

As in the VoD case, the final encodings selected to the sub-pilot may change, but will be probably be the same as already has been used up to now in the testing phase. The following is an example of codec settings of the contents tested which work correctly:

- Video → 3Mbps, 720\*576 (16:9), 25fps, AVC (High@14.0), GOP size = 960ms
- Audios → 128 kbps, 48 KHz, 2 channels, AAC (LC) (in 4 components: CAT (Catalan, dubbed), AD, CA, ENG (English, original language))
- MPD: time based.
- Chunks: 1.92s - 2.88s, 2\*GOP size - 3\*GOP size

### 8.3.Sub-pilots in Spain (Catalonia) – reception and user application

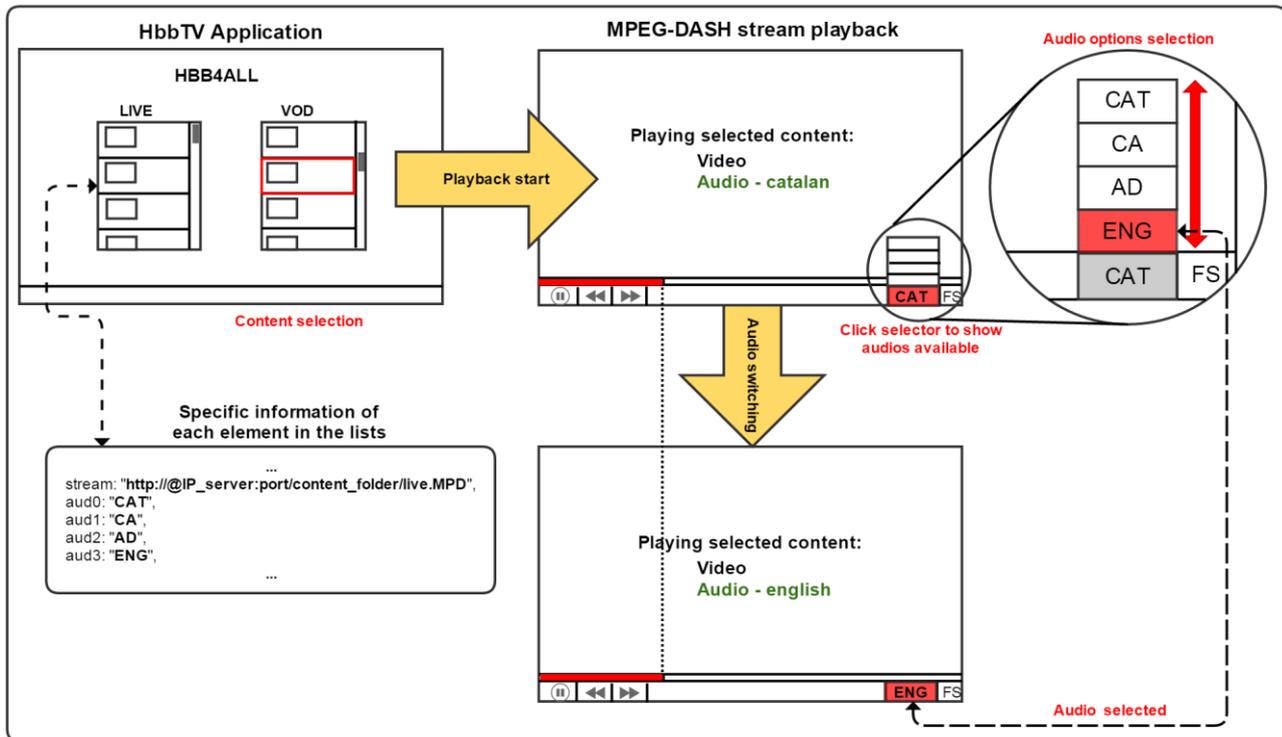


Figure 12. Reception and User Application

Figure 12 shows a picture of the reception and user application.

To differentiate if the streams the user can request are ‘Live’ or ‘VoD’, two separate lists of content are displayed in the test application, whose positions point to different available contents and examples (see Figure 12 top left). Each position of both lists contains all the necessary information to correctly playback the MPEG-DASH stream and to be able to switch between its different audio options, among other informational one like titles and content descriptions. The basic information is the absolute URL referencing a specific MPD hosted on the MPEG-DASH media delivery server, and the different language codes of the audios available, in the same order as they appear in the collection of audios of the stream. These last are used if the audio component languages cannot be retrieved in the HbbTV device, as explained in section 2.4.

The user can navigate with the TV remote control among all the positions in the lists and select the content which he prefers to play and test. When one is selected, a player in full-screen mode appears and the playback of the stream starts. The URL is used by the client device to get the MPD and start requesting the chunks of the video and the chunks of one of the audios by default. The player of the application is similar for the VoD and Live streams, both have buttons for going full-screen mode and back again and for the HbbTV audio options selector component; the VoD files additionally have play/pause, content back and forward buttons, plus a progress bar functionality.

The audio options selector is included in the player bar for all the cases in order to test the different audio services which are the target of the sub-pilots: CA, AD and other languages (see Figure 12 top right). The user can navigate to the selector, click on it to display the different audio options and select another audio to

change the audio being played. Then, when the new audio chunks arrive to the user device the player continues the video playback with the new audio, from the point of time where the audio was selected. This way, the users may test the quality of each additional audio service and compare it to the original audio.

For the audio options selector to work, it is required that the MPD and chunks of the contents are generated as a single MPEG-DASH asset containing the different audio components and the video, as explained in section 2.5.