D4.4

Version	1.0
Date	2016/12/23
Author	TVC
Dissemination status	PU
Document reference	D4.4

D4.4 – Pilot-B Evaluations and recommendations

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

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

Project no. 621014
HBB4ALL
Hybrid Broadcast Broadband for All

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

D4.4 – Pilot-B Evaluations and recommendations

Due date of deliverable: 2016/11/30
Actual submission date: 2016/12/23

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

Lead beneficiary for this deliverable: Televisió de Catalunya (TVC)

Revision 1.00

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

History Chart

Issue	Date	Changed page(s)	Cause of change	Implemented by
0.1	2016/09/30	all	First round of contributions included	TVC
0.2	2016/12/16	all	Second round of contributions included	IRT, RBB, TVC, UAB
0.3	2016/12/20	all	Fixing details from internal review	UAB
0.5	2016/12/22	all	Final version	TVC
1.0	2016/12/23	all	Format	UAB

Validation

No.	Action	Beneficiary	Date
1	Prepared	TVC	2016/11/25
2	Approved	UAB	2016/12/20
3	Released	UAB	2016/12/23

Table of Contents

Figures

Tables

1. 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 results from the service sub-pilots carried out during the operational phase of the project, in accordance with the Pilot-B objectives (Clean Audio, Audio Description, other languages and language learning). In these sub-pilots the developed services were tested, and qualitative and quantitative feedback was gathered from users from the mentioned target groups. Additionally, complementary user tests have been carried out, mainly regarding potential improvements for AD services and as preparation for the CA field trial.

The **Clean Audio sub-pilot in Germany (Berlin-Brandenburg)** ran for six weeks from May until July 2016 and consisted of a Clean Audio service offered via RBB's public Video on Demand service ("Mediathek"). Clean Audio versions of 4 RBB-programmes were published successively during the sub-pilot period for evaluation by a representative group of users. The sub-pilot aimed at verifying the effect of the Clean Audio processing under real life conditions by users with varying levels of hearing impairment. The sub-pilot was carried out cooperatively by RBB and IRT.

In preparation of the field trial, an additional lab test was carried out to verify the benefit of the Clean Audio generator for stereo signals, to have a good starting position with respect to the creation of audio material for the CA sub-pilots. Former lab tests had not yet produced conclusive results for stereo input material. IRT, in close cooperation with UAB, developed a new testing methodology for this work, to eliminate the weak points of the testing methodology that had been used up till then.

The field trial confirmed, that the CA generator can achieve small improvements, but also confirmed that the effect of the CA processing strongly depends on the content and the audio mix. The lab tests, being carried out in a much more controlled environment, showed a clear improvement for stereo material in speech intelligibility amongst the hearing impaired participants who do not wear a hearing aid. The results for 5.1 content had already shown (during the former project phase) that a significant improvement in terms of speech intelligibility can be reached with the developed CA generator.

In the **Clean Audio sub-pilot in Spain (Catalonia)** was carried out in September / October 2016 with the objective of – after successful lab tests in Germany – evaluating the benefit of Clean Audio in another language and culture. The sub-pilot was cooperatively carried out by UAB, IRT and TVC. The tests conducted in Catalonia also used stereo material and applied the newly developed testing methodology to verify the effect of Clean Audio on speech intelligibility. In addition to tests with hard of hearing, also tests with normal hearing people were carried out, to determine if their listening experience may be improved by applying Clean Audio, or at least is not worsened by it.

The tests confirm the results from the German sub-pilot. Tests with the hard of hearing without hearing aids indicate that the speech intelligibility for Clean Audio is better, when compared to the original audio mix. At the same time, normal hearing test participants indicate, that they do not consider Clean Audio as a disturbing factor, but on the contrary may benefit also with respect to the speech intelligibility.

The **Audio Description sub-pilot in Spain (Catalonia)** and the **Other languages and language learning sub-pilot in Spain (Catalonia)** were carried out by TVC as a combined service offering: the functionality of both alternative audio services was offered to the end user via TVC's public HbbTV Catch-Up TV service "TV3alacarta". TVC aimed at providing the Audio Description and Original Sound Track (OST) audio options through TV3alacarta (previously these features were only available via TVCs broadcast service), verify the enhanced workflow and evaluate the usage. The sub-pilot ran from April – September 2016. Unfortunately, due to technical reasons, the user data from the months August and September could not be used. To compensate for that, usage data from October and November were used in the evaluation.

Both the Audio Description and Original Sound Track deployments by TVC have successfully been added as new features in "TV3alacarta". Both features during the sub-pilots showed a stable acceptance. Based on the results, TVC has decided to maintain the AD and OST services as part of its HbbTV VoD service beyond the HBB4ALL project. Also, TVC is aiming at expanding this service and to bring these accessibility features to all platforms where TVC has presence. Large problems were encountered regarding the content rights, with respect to the distribution of OST versions via HbbTV VoD, currently resulting in a small amount of TV programmes that can be offered by TVC with this feature.

In summary, the objectives of Pilot-B (T4.3 and T4.4) have been achieved. The Clean Audio implementation, being a novelty as no actual service is on air at the moment, showed potential improvement of speech intelligibility both in lab and field tests in two different countries, languages and cultures. Further recommendations for the realisation of Audio Description services were provided, based on user tests, specifically regarding the sound editing and mixing. Last but not least, TVC maintains the provisioning of Audio Description as well as Original Sound Track on its VoD HbbTV service, the latter also being in the interest of normal hearing people with interest in language learning or watching programs in their native language.

2. Introduction

The Hybrid Broadcast Broadband for All project (HBB4ALL) has investigated Access 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. A new additional challenge is to offer viewers the opportunity to customise the Access Services they are using to best meet their personal preferences or needs.

HBB4ALL has tested 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. During the operational phase of the HBB4ALL project (for all Pilots A to D running from August 2015 – July 2016) the project partners implemented field tests to gather user feedback and to assess the acceptance and quality of services in various scenarios. For these tests, a number of different so-called sub-pilots had been scheduled to be carried out in the operational phase.

In the specific case of Pilot-B (Work Package 4 in the project), novel audio services have been tested and partly also prototypically rolled out in the different HBB4ALL countries to address the above challenges and to receive and evaluate user feedback for later improvement of the services. In parallel to the field tests, also complimentary qualitative lab tests were carried out as part of the sub-pilots.

2.1. Purpose of the Document

HBB4ALL deliverables "D4.1 – Pilot-B Progress Report" **[1]** and "D4.2 – Pilot-B Solution Integration and Trials" **[2]** provided an overview of the Pilot-B activities and achievements during the first 20 months of the project timeline. The preparations for the sub-pilots, specifically including the technical developments and preliminary user (lab) tests and their outcome, were described there. The reader is kindly referred to these documents for detailed information (specifically regarding the implementations of technical components by the partners).

The current document – being the final deliverable for Pilot-B – gives an overview of all sub-pilots carried out in Pilot-B during the operational phase of HBB4ALL. For each sub-pilot, chapters 3, 4, 5 and 6 respectively give an overview of the goals, the piloted service, how it was set up, what was tested and for how long, which user groups were targeted and which testing and evaluation methodology was used. Last but not least, the evaluation of the sub-pilots' outcome, the results and recommendations resulting from the sub-pilots are provided.

In addition to the sub-pilots, chapter 7 documents the complimentary user tests that have been carried out since the submission of deliverable D4.2 (30/09/2015) **[2]** , to complement the user tests that had been described there.

2.2. Organisation of Pilot Phase

Whereas all sub-pilots in Pilot-B addressed audio services, they still were very different in nature, as they pursued different objectives. This resulted partly from the difference in the maturity of the technical solutions that were used, partly from the difference in organisational possibilities and facilities available to the partners that were involved. Also the timing of each sub-pilot was adapted to the available resources as

well as to the respective goals (for details please refer to the descriptions in the further chapters). **Table 1** gives an overview of the sub-pilots that have been carried out during the operational phase of Pilot-B and of their respective objectives.

Table 1. Overview of sub-pilots in Pilot-B

Sub-pilot name	**Sub-pilot objective**	**Partner(s) involved**
Clean Audio – Sub-pilot in Germany (Berlin-Brandenburg)	After successful lab tests, evaluate the Clean Audio tool in a field test under real life conditions.	*RBB*, IRT
Clean Audio Sub-pilot in Spain (Catalonia)	After successful lab tests in Germany, carry out further user tests to evaluate the benefit of Clean Audio in another language and culture.	*UAB*, TVC, IRT
Audio Description Sub-pilot in Spain (Catalonia)	Verify the enhanced Audio Description workflow and the adapted HbbTV application as part of an operational service and evaluate its usage.	*TVC*
Other languages Sub-pilot in Spain (Catalonia)	Verify the enhanced workflow for inclusion of Original Sound Track audio and the adapted HbbTV application as part of an operational service and evaluate its usage.	*TVC*

During the first 20 months of the project (tasks 4.1 and 4.2) technical and organisational preparations for the respective sub-pilots had been carried out. At the start of the operational phase, these preparations were finalised. For each sub-pilot a separate time schedule was defined, depending on e.g. the remaining work required to start the sub-pilot (organisational, technical implementations) as well as alignment of the timing with activities for sub-pilots in other HBB4ALL pilots (specifically Pilot-A where many of the partners involved in Pilot-B also participated and contributed to sub-pilot activities).

Each sub-pilot was coordinated by a dedicated partner (shown *italic* in **Table 1**). In case of the Clean Audio sub-pilots, they worked in close cooperation with other partners. Amongst them, the activities required to prepare and carry out the sub-pilot were distributed. For each sub-pilot an evaluation methodology was agreed on; due to the different nature and objectives of the sub-pilots, it was not possible in Pilot-B to use an overall valid method for all sub-pilots.

2.3. 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.

2.4.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 caches 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.

Original Sound Track (OST) – Is the Original Sound Track assigned to the video when it was initially produced.

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[1].

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.

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

3. Clean Audio – Sub-pilot in Germany (Berlin-Brandenburg)

From previous projects such as DTV4ALL we know that for viewers with hearing impairments Clean Audio can increase understanding of dialogue in TV shows by reducing the level of noise, music, atmosphere or any other signal which is not speech to a minimum. In HBB4ALL, IRT developed a process for the automatic creation of a Clean Audio stream. Using samples of television content, the sound was processed and the results tested in numerous lab tests. This work is documented in HBB4ALLdeliverables 4.1 **[1]** and 4.2**[2]**. The following describes the Clean Audio sub-pilot in Germany conducted by RBB in cooperation with IRT.

3.1. Goals of Sub-pilot

Results from three Clean Audio lab tests conducted by IRT and RBB in HBB4ALL indicated that the Clean Audio processing tool developed by IRT had a positive effect on the understanding of speech in TV programs. In a lab environment, testers with varying levels of hearing impairment were shown a series of short clips from TV shows. The original version and Clean Audio versions of the clips were shown. The test participants rated their understanding and experience of each clip.

The aim of this sub-pilot was to investigate in a field trial if similar results could be achieved when Clean Audio processing was applied to a complete TV programme and the programme was viewed under real life conditions by users with varying levels of hearing impairment.

3.2. Description of Service / Application

3.2.1. Technical Implementation

RBB and IRT worked together to produce the Clean Audio videos used in the sub-pilot. The content used for the tests originated from RBB's "Mediathek" VoD service. The audio files were exported from RBB's VoD production environment and handed over to IRT. The Clean Audio processing itself was completed by IRT. RBB then integrated the Clean Audio files back into its online VoD production system. The Clean Audio versions of the videos were then published in the regular RBB Mediathek.

3.2.2. Functionalities

The videos used in the Clean Audio pilot were available on demand in the standard RBB and ARD "Mediathek" Catch-Up TV service for the duration of the sub-pilot. For the purpose of the sub-pilot four TV shows were selected, which were available in the original version as well as a Clean Audio version. The Clean Audio version was clearly labelled: this included a graphic indication on the thumbnail image and the phrase "Test Clean Audio" in the video title. When the video was selected from the overview page, the preview page contained information about the sub-pilot.

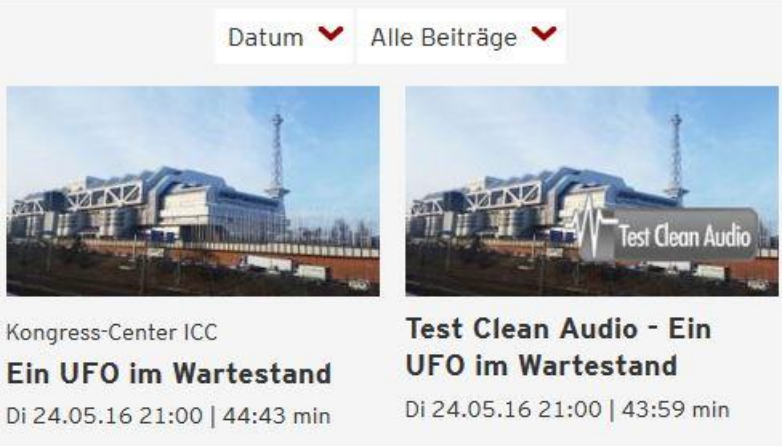


Figure 1. Teaser for original and teaser for Clean Audio version of programme in RBB Mediathek

3.2.3. Availability of Service

Clean Audio videos were available in the RBB Catch-Up TV service RBB Mediathek and the ARD Mediathek. The central ARD Mediathek collects and offers all Mediathek content from ARD regional broadcasters. Accordingly, videos published in the RBB Mediathek are automatically also included in the ARD Mediathek. For each Mediathek there is an HbbTV version of the service, an online PC optimised version and a mobile version.

Clean Audio versions of programmes were published successively on a weekly basis over a period of four weeks. The first one was published on Monday, 20th June 2016. All videos were de-published on 18. July 2016.

- 20.06.2016: Nerven und Nerven lassen - Was uns auf die Palme bringt
- 27.06.2016: Landschleicher Extra. Ausflüge in die Mark Brandenburg
- 04.07.2016: Ein UFO im Wartestand - Kongress-Center ICC
- 11.07.2016: Sanssouci von oben

During this time a recruited group of testers were requested to watch all four programmes and complete an online questionnaire. As the videos were available in the regular Mediathek, any user could watch them. Each Clean Audio video was clearly labelled as such – this included a label on the video thumbnail (see **Figure 1**), a short explanation of Clean Audio and a link to further information about the test. The webpage of RBB Innovationsprojekte[2] included a link to a simplified questionnaire for viewers who wished to take part in the test or provide feedback.

3.2.4. Intended Audience

Clean Audio versions of TV programmes are expected to offer a benefit to TV viewers, by producing an audio version with improved speech intelligibility. As such it is specifically aimed at viewers who are hard of hearing, both those with and without hearing aids.

[2] http://www.rbb-online.de/unternehmen/der_rbb/profil/zukunft/die_innovationsprojekte_0.html

3.2.5. Workflow / Production Aspects

The content used for the sub-pilot was produced using the regular RBB "Mediathek" VoD service production environment. The RBB VoD production environment stores the so called master file which works as base for all renderings and contains separate audio tracks for each stereo channel. Once the videos to be used in the sub-pilot were identified, the following workflow was followed:

1. Masterfile download from Adobe CQ5 which is originated by Arvato VPMS
2. Video split using MP4Box, result: WAV Files for each stereo channel
3. FTP transfer from RBB to IRT (just the extracted audio)
4. Clean Audio processing by IRT
5. FTP transfer from IRT to RBB (the Clean Audio enhanced audio)
6. Re-join of video & audio using FFMPEG
7. Upload to RBB VOPS (Video On3.line Production System)
8. VOPS processing to create new video assets
9. Adobe CQ5 - Duplication of necessary documents to replicate a RBB broadcasting schedule
10. Adobe CQ5 - Assignment of the new video assets to the replicated broadcasting schedule documents
11. Timed activation
12. Transport via ARD internal interface for video asset metadata an activation
13. Import to ARD Mediathek, the central ARD media library (RBB media library is a part of this system)
14. Automatic activation within the RBB & ARD Mediathek
15. Deactivation time based after pilot expiration

The above workflow was adhered to by RBB and IRT for the duration of the pilot.

3.3. Description of User Tests

3.3.1. Aim

The aim of the field test was to determine if understanding of dialogue for hard of hearing viewers improves when watching Clean Audio versions of programmes under real life conditions. A further aim was to see if Clean Audio processing created any other, unexpected results.

3.3.2. Methodology

To achieve the aim of the test a pragmatic methodology was developed and applied. First, four programmes were selected for the test. The selection of the videos was based on an analysis of complaints made to RBB's service hotline relating to the sound quality of TV programmes broadcast by RBB. To be suitable for the test the programmes had to be available in the Mediathek for the duration of the field test. RBB needed to have rights' clearance to use the programmes for the test. The ideal duration of the videos was aimed to be approximately 30 minutes.

A test group of 9 hard of hearing testers helped to identify positions in the original programme where it was difficult for hard of hearing viewers to understand the dialogue. They were asked to watch the original programmes in the RBB Mediathek, under their individual normal viewing conditions. This included using the consumer devices they would typically use for Mediathek content, i.e. Smart TV, PC, sound bar, etc. and using a hearing aid if they normally used one when watching TV. They were asked to note any problems they had understanding the dialogue while watching each programme. For each problem they should also note the timecode at which it occurred. The testers later entered this information into an online questionnaire. The results of this test were gathered to identify the "problematic" scenes in each programme (with respect to dialogue intelligibility).

For each programme in the test, a Clean Audio version was produced and published in the Mediathek. A group of 26 hard of hearing testers helped identify positions in the Clean Audio video where it was difficult for them to understand the dialogue. This was done by replicating the above test. The results of this test again were gathered to identify the "problematic" scenes in each programme.

The results of both tests were then compared to find out if there were less problematic scenes after Clean Audio processing, if there were any unexpected problems after Clean Audio processing and how Clean Audio processing affected the overall rating of the audio in the videos.

3.3.3. Testers

Recruitment

For the tests we originally aimed to recruit 30 hard of hearing testers who were familiar with and had access to the RBB Mediathek. The group should be as representative as possible, including men and women of different ages, half of whom wore a hearing aid and half without. The testers were required to be at least 18 years old.

Testers were recruited via several channels. The local user associations were provided with information about the tests and asked to inform their members. A number of testers were recruited via the RBB tester database. This is a database with names and contact details of testers from previous tests. RBB and IRT worked together to recruit further testers to whom IRT had contact. Details of the test and a call for testers were also published on the RBB website (www.rbb-online.de/innovationsprojekte).

Testers

The final test group recruited consisted of 35 people, of whom 18 used a hearing aid and 17 did not.

Table 2. Clean Audio test group constellation

Age	Percentage of total population	Number of testers
18-24	5,7%	2 candidates
25-39	8,6%	3 candidates
40-59	31,4%	11 candidates
60 and older	54,3%	19 candidates
male/ female	51,4 / 48,6%	18 / 17 candidates
Berlin/ Brandenburg	28,6 / 22,8%	10 / 8 candidates
Other regions	48,6%	17 candidates

All testers who actively participated in the test received a compensation of 100 Euros.

3.3.4. Report on Test

The test was run in two stages. For the first stage ("pre-test") a group of 10 testers were recruited to evaluate the original soundtrack of the TV shows identified. From the testers who responded to the call to participate ten were selected. Communication with the testers was primarily conducted by email. In May 2016 (21st calendar week), the testers were sent an email with the list of TV shows to watch, instructions of how to find them in the Mediathek and a link to the online questionnaire. A form was attached to the email, testers could print this and use it to note down the times and problems while watching the videos. Of the recruited testers, nine participated and provided feedback on all four programmes.

For the second stage of the test (the "field trial") 30 testers were recruited. These testers were asked to watch four Clean Audio videos over a period of four weeks from mid-June (25th calendar week) until mid-July 2016. The testers were sent a mail on Mondays with the name of the video, instructions of how to find it in the RBB Catch-Up TV Mediathek and a link to the online questionnaire. A form was attached to the email, testers could print this and use it to note down the times and problems while watching the video. In total 26 testers completed all four questionnaires. Communication with the testers was primarily by email.

The testers appeared to understand the principle of the test as there were very few queries or questions and a high rate of participation. The quality and quantity of information returned varied, as is to be expected with a group of that size.

One of the videos identified for inclusion in the test had to be removed from the Mediathek for a short time due to a legal dispute, not related to the project. The issue was solved quickly and the video re-published. It had no effect on the test.

3.4.Evaluation of Sub-pilot

Both RBB and IRT participated in the evaluation of the sub-pilot. RBB collected and documented the results; the resulting data were then provided to IRT for further analysis. The evaluation of the analysed data again was a joint activity.

3.4.1. Questionnaire

To gather feedback from the testers an online questionnaire was created using the online survey platform "Survey Monkey". (see Annex **11.1**, questionnaire in German).

The questionnaire was divided into three sections. The first section gathered personal details relevant for the analysis of the feedback such as age, gender, degree of hearing loss, if the testers wore a hearing aid and on the hardware used to view the videos and playback the audio.

The second section of the questionnaire was designed to collect feedback on the scenes where testers had problems understanding the audio. For each problem the tester was asked to enter the timecode in the video and describe the problem in their own words. The approach was chosen to avoid any undue influence on the testers.

The final part of the questionnaire asked the tester to rate the audio and also left room for them to add any other comments.

The testers were sent an email each test week with a link to the questionnaire. Attached to the email was a printable sheet on which the user could take notes about the problematic scenes while watching the videos. This way they could then enter them into the online questionnaire at a later date.

While RBB knew testers' email addresses, the replies gathered by the online survey platform were anonymous in order to fulfil the ethical test requirements.

3.4.2. Evaluation

3.4.2.1. Data Analysis

The results of the second section in the online questionnaire (see Annex **11.1**) were lists of timecodes (at which problems with the intelligibility were found) and corresponding comments for each of the four programmes under test. The first step was to sort these lists of problems according to the timecode. The resulting list of problems (timecodes with bad intelligibility) was transformed into another list, in which the overall number of reported problems (for all test participants) was counted for every second of the programme. For example, an entry for bad intelligibility at 1, 5 and 9 seconds and two entries at 4 seconds results in a list as shown in the **Table 3** below.

Table 3. Results for bad intelligibility

Timecode	00:00	00:01	00:02	00:03	00:04	00:05	00:06	00:07	00:08	00:09
Problem count	0	1	0	0	2	1	0	0	0	1

The resulting list of these problem counts for one the programmes under test ("Landschleicher Extra. Ausflüge in die Mark Brandenburg") is shown graphically as an example in **Figure 2**.

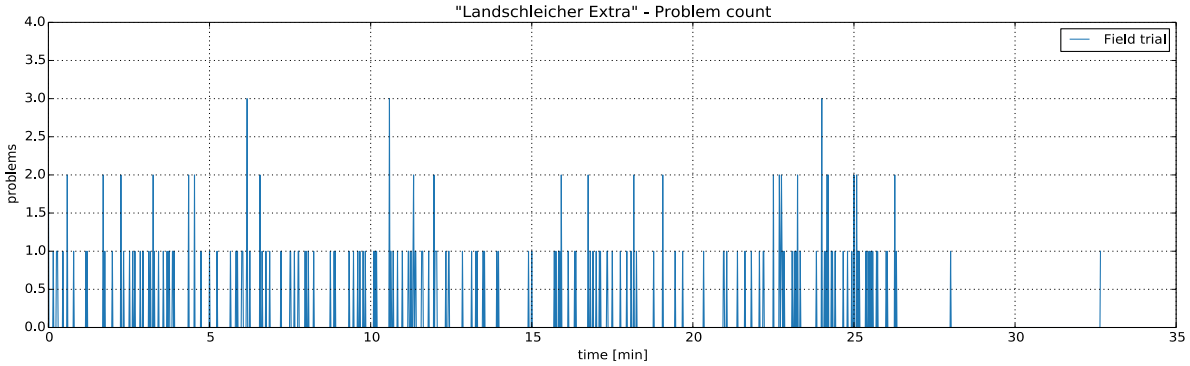


Figure 2. Amount of problem entries counted for every second of the programme under test

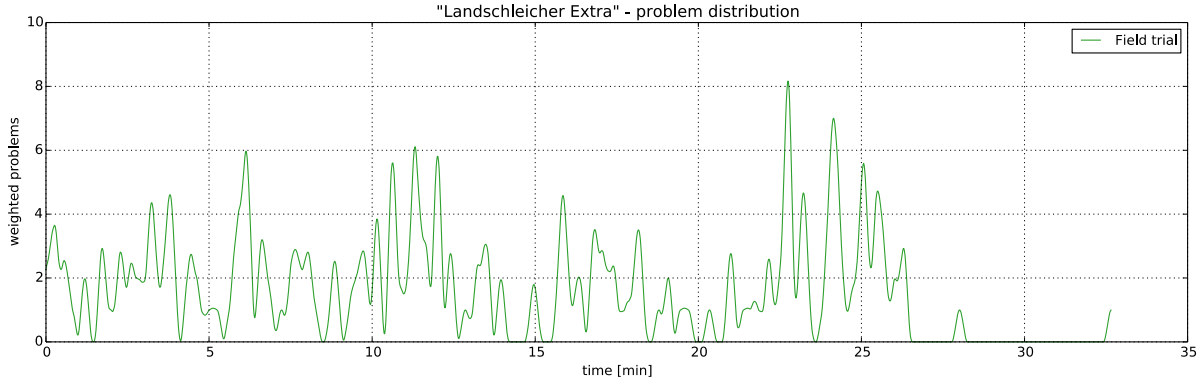


Figure 3. Weighted problem distribution of the programme under test

The graph in **Figure 2** shows that most timecodes were only mentioned once (1.0 on the y-axis) as a problematic point in the programme by any participant in the test. However, typically bad intelligibility usually does not just occur at single time points but rather during periods of a few seconds. To find these periods with adjacent problem entries we looked at every second in the list, combined the problems within a window of 30 seconds[3] around this time code and weighted them based on their distance from the time code under consideration. The result of this weighted summation – the weighted problem distribution – is plotted as a graph, shown as an example in **Figure 3.**

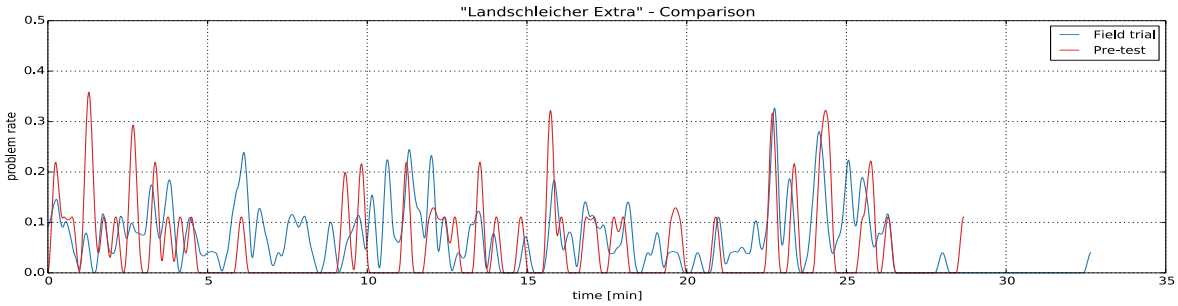


Figure 4. Comparison of the normalised problem distributions of the field trial and the pre-test

[3] Several window-sizes between 10 and 50 seconds were tested. A size of 30 seconds showed the best results.

To compare the results of the field test and the ones from the pre-test, the results needed to be normalized. This was done by dividing the values, which were generated by the weighted summation, by the number of participants for the respective test, resulting in the so-called problem rates. The resulting graph is shown in **Figure 4**.

In this graph the problem distributions from the field test (blue) and from the pre-test (red) are compared to each other. For each tested programme (see the evaluation in the next subsection) it can be seen, that sometimes the problem rate is higher for the pre-test results and sometimes for the results from the field test.

3.4.2.2. Evaluation

The following is an evaluation of the user tests conducted in the German Clean Audio sub-pilot. The evaluation is based on the data analysis conducted by IRT. For each TV show used in the user test the problems identified in the pre-tests and field tests have been mapped on each other. The positions in the videos where there was a significant/obvious high problem rate for either or both groups were identified. The aim is to gain a more detailed understanding of the problematic positions and if clean audio (CA) processing leads to improved intelligibly. Each position has been evaluated in terms of factors affecting the audio and user comments.

Landschleicher Extra

Landschleicher is a short weekly report on RBB TV. Every week "Landschleicher" portrays a location in the federal state of Brandenburg with less than 2000 inhabitants, telling stories about the area and the people who live there. There are lots of interviews with the local inhabitants and footage of them going about their daily business. The Landschleicher Extra used in the test is a collection of several individual Landschleicher reports.

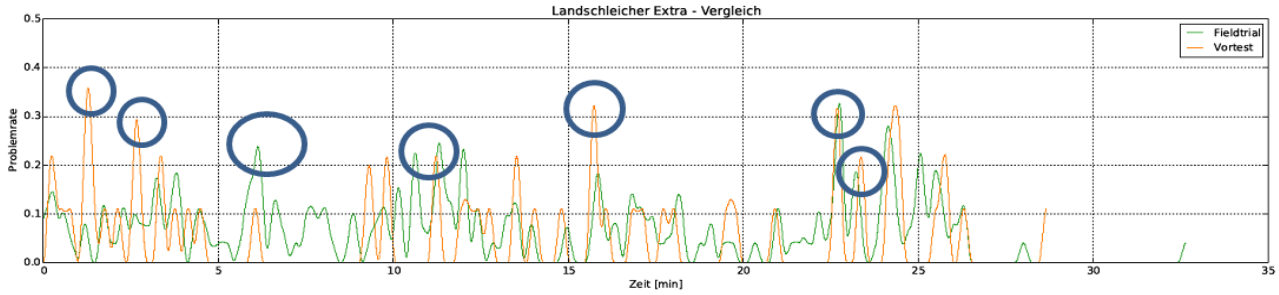


Figure 5. Normalised problem distributions of the field test and the pre-test for the video "Landschleicher Extra. Ausflüge in die Mark Brandenburg"

Table 4. Evaluation "Landschleicher Extra"

Timecode*	Problem situation**	Sound setting***	Evaluation, comments
01:19	PT	- An off-camera speaker introduces location - There is loud background music - There are additional photo trigger	This is a scene where CA appears to have made difference

		sounds between shots	
02:41	PT	- The scene is filmed in a church with a lot of echo - The interview partner has a tendency to have a sibilant voice	This is a scene where CA appears to have made difference
06:00 – 09:00	FT (only FT problems reported)	- An artist is interviewed in his studio - The microphone position seems to be slightly unsteady -The speakers are generally not speaking in the direction of the camera	No obvious explanation why the problem rate in field test is slightly higher than in the pre-test. Most of the problems reported during the field test appear to relate to artist's speech being too quiet or unclear.
11:15 – 11:20	=	- This is an interview situation and with people speaking with a strong dialect - There is an amount of echo in the room	As the main problems appear to be dialect, there was a very similar problem rate in both groups. CA processing did not improve intelligibility
15:45	PT	- This is an open air interview situation - There is chainsaw background noise while the off-camera speaker is speaking	This is a scene where CA appears to have made difference
22:42	=	- This an open air / street scene with an interview with a local - The interviewee speaks with a strong local dialect	As the main problems appear to be dialect related CA processing did not improve intelligibility
23:20	=	- This scene features a group of interviewees on the street - They are all are talking at the same time - The reporter is louder than the interviewees	CA processing did not appear to make any difference to the intelligibility of this scene The scene itself seems to be more an editorial introduction to the next scene without any requirement of speech intelligibility

Nerven und Nerven lassen

The documentary "Nerven und Nerven lassen – Was uns auf die Palme bringt" (To annoy and be annoyed – what drives us crazy) is about everyday annoying situations that appear to have an increasing impact on our life. The documentary is produced in a style that is intended to be annoying and reflect stress situations. To achieve this effect, the director uses aggressive background music. Generally, the video sound is quite flat.

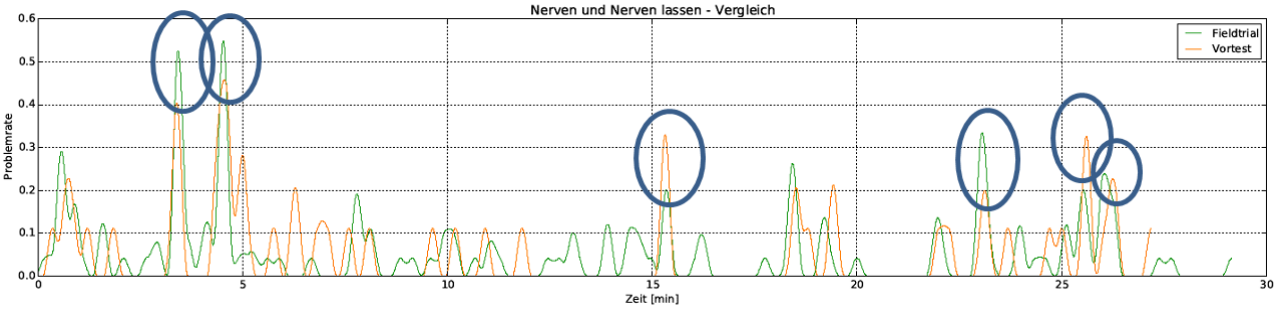


Figure 6. Normalised problem distributions of the field test and the pre-test for the video "Nerven und Nerven lassen - Was uns auf die Palme bringt"

Table 5. Evaluation "Nerven und Nerven lassen"

Timecode*	Problem situation**	Sound setting***	Evaluation, comments
03:24	=	- The scene portrays an annoying/stressful morning situation in a household - A woman speaks to her husband from behind a bathroom door - There is an extreme echo from the voice in bathroom	CA processing made no real difference in this scene. A high level of speech unintelligibility is to be expected due to the situation it is portraying.
04:34	=	- This scene is an interview with medical scientist in large hospital corridor - There is a lot of echo - The interview partner speaks in a low voice, similar to a whisper due to the hospital environment - There are additional music effects in the background which transport the topic „stress"	CA processing made no real difference in this scene. This is a scene in which several factors contribute to the low speech intelligibility
15:20	=	- This scene is and interview with a	CA processing made no real difference in this scene. This is a scene in which several

		psychologist. - There is background music that seems quite close to the pitch of the psychologist's voice. The psychologist's does not always speak clearly, certain word are "swallowed"	factors contribute to the low speech intelligibility
23:05	=	- This scene is an interview setup - There is a mix of home & street scenes - The interview partner mumbles somewhat - There is background music	CA processing made no real difference in this scene. This is a scene in which several factors contribute to the low speech intelligibility
25:37	=	- In this scene there is a lot of whispering & yelling and the sound is distorted. - The scene is designed to be theatrical to illustrate stress levels in job review situations.	CA processing made no real difference in this scene. This is a scene in which the sound is purposely distorted
26:15	=	Note: artistic effect to simulate stress situation. Sound purposely distorted	CA processing made no real difference in this scene. This is a scene in which the sound is purposely distorted

Video: Sanssouci von oben

The video is about the UNESCO world heritage site Park Sanssouci. It is filmed from the air and for most of the film an off-camera speaker describes the park and palaces. In two scenes a park representative speaks directly into the camera. The off-camera speaker is the well-known German actress Katharina Thalbach. Thalbach has an easily recognisable voice that some testers found irritating.

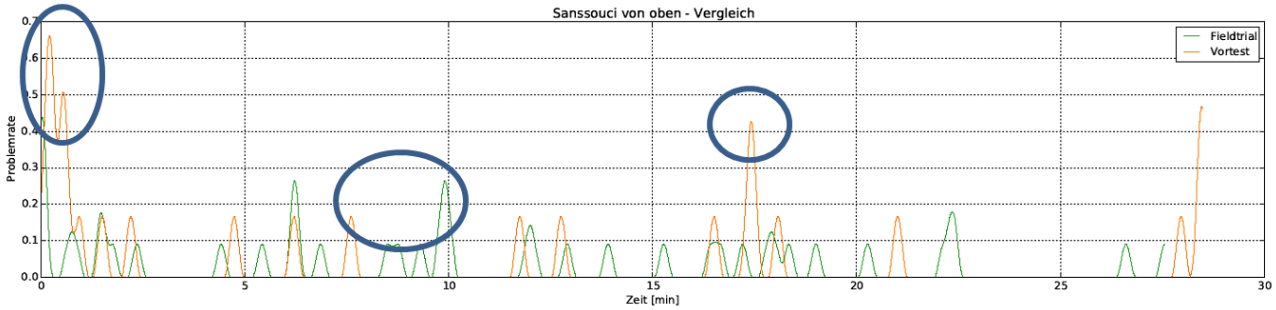


Figure 7. Normalised problem distributions of the field test and the pre-test for the video " Sanssouci von oben"

Table 6. Evaluation "Sanssouci von oben"

Timecode*	Problem situation**	Sound setting***	Evaluation, comments
00:00 – 00:40	PT	- This is the opening scene with relatively loud background music - The off-camera speaker starts to speak in soft voice and with a slight lisp.	This is a scene where CA appears to have made difference.
08:00 – 10:30	FT (only FT problems reported)	- In this section has long spoken passages and changing background music - The speakers voice often drops quite low and is hard to differentiate from background music	This is a scene where CA appears not to have made a difference. In fact there were was a higher problem rate recorded after the CA processing. There is no obvious explanation for this, but the off-camera speaker's voice irritated a lot of testers, additionally there is not much contrast between the speaker's voice and the music.
17:26	PT	In this section has long spoken passages and changing background music - The speakers voice occasionally drops quite low and is hard to differentiate from background music - At this point in the film the speaker uses a foreign name "Belvedere", the name of the palace	This is a scene where CA appears to have made difference. But the use of a foreign term spoken in a soft voice appears to have caused intelligibly problems.

Video: Ein UFO in Wartestand

This video is a documentary about the International Congress Centre in Berlin which is situated on a very busy junction with lots of traffic. It contains a mix of archive footage and lots of interview some of which were filmed inside or in front of the building.

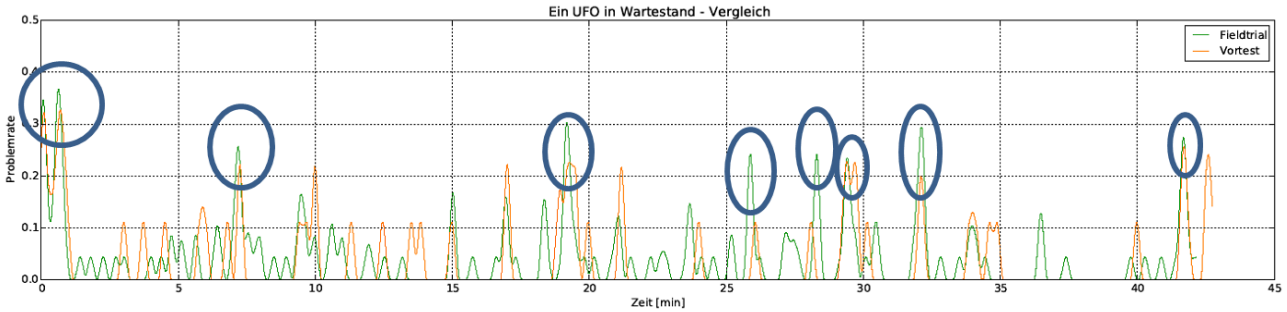


Figure 8. Normalised problem distributions of the field test and the pre-test for the video "Ein UFO in Wartestand"

Table 7. Evaluation "Ein UFO in Wartestand"

Timecode*	Problem situation**	Sound setting***	Evaluation, comments
00:00 – 00:50	=	- The first scene in the documentary starts with an interview with the architect inside the empty building - The architect, Ursulina Schüler-Witte is now over eighty years old, her voice is relatively quiet and appears to have weakened with age. - There is dramatic background music	This is a scene where CA appears to have made no difference to the intelligibility. Testers in both tests had problems understanding the architect.
07:11	=	- The architect is speaking and there is a lot of very loud traffic noise in the background. - The traffic noise was added later.	This is a scene where CA appears to have made no difference to the intelligibility. The problem was the combination of street noise and the weak voice
19:11	=	- This scene is an interview with a male expert witness filmed in front of the building. He speaks with a lisp which means his pronunciation is somewhat unclear - There is very loud traffic noise in the background	Even after CA processing the background noise was too loud relative to the speaker for many testers. CA processing did not make any significant improvement as the traffic and speech were recorded together.

		- The traffic noise had originally been recorded together with the speech.	
25:53	FT	- This scene is an interview with a male interviewee - There is loud traffic noise in the background. The traffic noise was originally recorded together with the speech.	No obvious explanation why there was a higher problem rate in the field test. CA processing did not make any significant improvement as the traffic and speech were recorded together.
28:19	FT	- This scene is more of the interview with the male expert witness filmed in front of the building. - There is very loud traffic noise in the background. - The traffic noise was originally recorded together with the speech.	See above
29:24	=	- This scene is an interview with a female architect filmed in front of the building. - There is very loud traffic noise in the background	CA appears to have made no difference. Users found that the speaker spoke unclearly and too fast
32:06	=	- This scene is an interview with a male architect in his office. He is leaning over plans on a desk and talking. - His voice is quiet and pronunciation not always clear - There is no additional background music or noise.	Even after CA processing the architects voice was not strong and clear enough for most of the testers to understand him
41:42	=	- This scene is an interview with a male interviewee filmed in front of the building. - His speech is not always very clear - There is very loud traffic noise in the background	Even after CA processing the there was still too much background noise for most of the testers. CA processing did not make any significant improvement as the traffic and speech were recorded together.

		- The traffic noise was originally recorded together with the speech.	

* Either timecodes at which an eminent problem rate occurred, or a time interval with an eye-catching problem rate distribution

** Either noticeable differences between the problem rates in the pre-test (PT) and the field trial (FT), or eye-catching high problem rates for both (=)

*** Noticeable features of the sound mix or the speech, reported by listening to the video at the mentioned timecode

Findings

In the German sub-pilot CA processing appears to have been most effective, i.e. resulted in increase in speech intelligibility in scenes with loud background music or noise and a single speaker. However, scenes where the background noise was extremely loud relative to the speaker or deliberately distorted for artistic effect did not appear to benefit significantly from the CA processing used in the user tests. In cases where the background noise was originally recorded together with the speech, CA processing did not improve the intelligibility, as such noise cannot be separated from the speech retrospectively.

Once other factors such as unclear pronunciation, dialect or quick speech are also present then the effectiveness of CA processing is reduced. This is to be expected as CA processing cannot improve the effect on intelligibility caused by such factors.

Scenes were the speaker is not speaking directly in the microphone or is not visible appear not to benefit from CA processing in terms of speech intelligibility.

In conclusion, the biggest hindrances for the testers to understanding the dialogue in the video appear have been

- Speakers voice and/or speech (pitch, speed, pronunciation, dialect)
- The recording situation (background noise during recording, movement, speakers not speaking into microphone)
- The audio mix, for example effects introduced for artistic effects.

While CA processing improved the speech intelligibility results in some cases, the more complex the sound setting or the presence of more influencing factors seem to lead to fewer improvements.

Following side points should be considered with respect to the interpretation of the field test results and their validity in a wider scope:

- All programs used in the field trial were documentaries; no content from other genres was tested.
- The results strongly depend on the content; exactly for this reason the more detailed evaluation of the results for each program was provided.
- The pre-test, in which the unprocessed versions were evaluated, had a relatively small number of participants (9).
- All programs used in the field trial had stereo audio.

3.4.2.3. Conclusions

The CA pilot in Germany, carried out with limited resources, set out to test CA for the first time under real life conditions in users' home environments. From the pilot's results it can be seen, that the effect of Clean Audio processing strongly depends on the content and the audio mix. Of the four programs used in the pilot, one showed a clear improvement in speech intelligibility and one showed a small improvement. For the other two, the Clean Audio processing had no effect. However, it can also be seen that the speech intelligibility in general does not worsen by applying the Clean Audio method. All four programs were available with stereo audio mix (as input to the Clean Audio generator).

There is no clear difference in the speech intelligibility experienced in general by test participants with and without hearing aids. This could indicate, that the single version of Clean Audio used in the pilot could work equally well for these two user groups.

Regarding the test methodology it should be mentioned that, because only nine participants took part in the pre-test, the comparison of the field test results with the pre-test results should be taken with some precaution.

The effect of the various playback devices (e.g. TV with its own speakers, TV with soundbar, PC) on the speech intelligibility could not be measured based on the pilot results. The number of testers was not large enough for such an analysis.

The evaluation of the four individual programs confirms that speech intelligibility is highly influenced by the conditions under which audio recordings are made, as well as the speakers and by the audio mix. Speech intelligibility sometimes can already be improved by optimizing these, probably leading to less complaint by TV viewers.

3.5. Results and Insights

It is in all broadcasters' interest to offer their audience the best possible picture and audio quality. Following the implementation of the EBU guideline R128 on loudness normalisation and permitted maximum level of audio signals in 2012, ARD and ZDF turned their attention to the TV-viewers' satisfaction regarding audio quality. As a result, in 2014 recommendations were published for ZDF and all ARD members, covering the production, editing and audio mixing in addition to recommendations for the equipment and settings required to control the audio mix[4].

Despite these recommendations, feedback from the RBB service hotline showed a large number of viewers' complaints concerning the audio quality of TV programmes. When analysed, the complaints could be broadly divided into two categories: they were either speaker-related (poor enunciation or strong dialect) or they were related to the audio mix. While stricter adherence to the above mentioned recommendations could possibly solve some of the problems, there remains a need for an even clearer dialogue among a certain proportion of the audience.

As with other services a "one size fits all" solution does not appear to be the answer for speech intelligibility. The lab tests and field trail conducted in HBB4ALL indicated that the Clean Audio processing developed in the project can offer an improvement. The Clean Audio processed videos used in the pilot were publically

[4] "Sprachverständlichkeit im Fernsehen. Empfehlungen für Programm und Technik", July 2014 © ARD/ZDF (German only)

available in the RBB VoD service for the duration of the pilot and during the IFA and IBC trade fairs. The videos remain in the system but are no longer publically available. RBB would welcome further testing to better understand under which conditions and for what content Clean Audio is most suited for viewers. Further discussions with ARD members have also suggested testing Clean Audio generator as a method to improve machine understanding of dialogue.

4. Clean Audio Sub-pilot in Spain (Catalonia)

4.1. Goals of Sub-pilot

Based on lab-tests in Germany, UAB carried out lab-tests in Catalonia. The goal of this sub-pilot was to verify the benefits of Clean Audio in a different culture and for a different language.

4.2. Description of service / application

4.2.1. Technical implementation

In the following block diagram, the manual processing steps required to produce the final materials for the Clean Audio lab test are explained:

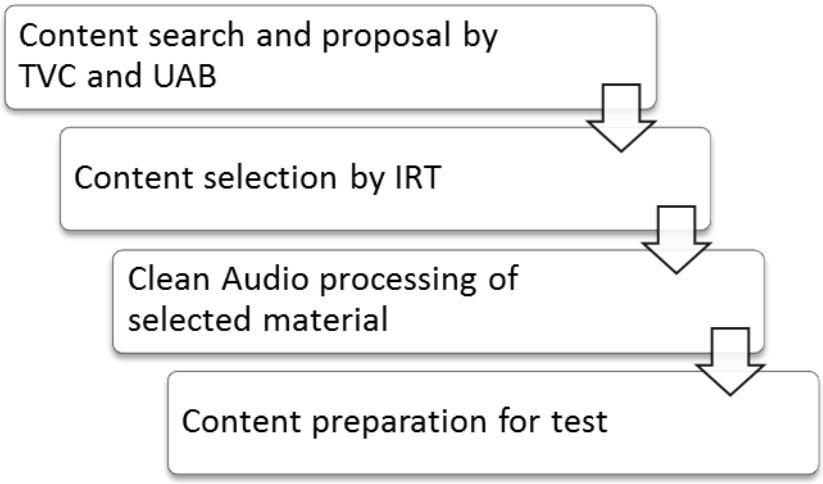


Figure 9. Clean Audio Catalonia sub-pilot workflow

To generate Clean Audio contents, UAB and TVC selected a number of clips from the TVC archive which presented an appropriate level of background sounds interfering with the intelligibility of the dialogue. From TVC's archive only audio visual material containing stereo audio was available. After this selection by TVC and UAB, IRT carried out a more thorough selection, since not all audio mix is suitable for subsequent Clean Audio processing. Once the final selection had been completed, IRT processed the audio of the selected material. Finally, the materials were remultiplexed and put it in the test environment (hosted by IRT). For further technical details on the Clean Audio generator, please refer to D4.2 **[2]**.

4.2.2. Functionalities

In the lab test, four different audio versions were compared with respect to speech intelligibility: the original audio and three Clean Audio processed versions, A, B and C. For further details on the different types of processing, please refer to D4.2 **[2]** and section **11.2** of this deliverable.

4.2.3. Availability of Service

This sub-pilot was executed as a closed lab test, so it was not implemented as an open service. TVC has followed this activity and will verify the results of the sub-pilot, as a basis for a decision on the possible provision of a Clean Audio service to its VoD offer.

4.2.4. Intended Audience

Clean Audio aims to improve speech intelligibility, and its intended audience is hard of hearing users, for example people who have limited frequency hearing. However, frequently hard of hearing people and normal hearing people watch television together. For this reason it was important to test Clean Audio with mixed hearing audiences who may use such a service.

This lab test aimed to prove the effectiveness of Clean Audio for the following groups of users:

- The hard of hearing
  Not all hard of hearing members of the community are able to enjoy reading subtitles, therefore they require a louder dialogue than people with full hearing capacity. However, when users reach for the remote and raise the volume, they are raising the volume of the whole audio mix. This may result in a lack of intelligibility, due to all the background sound effects, music and noises being louder, despite the fact that the hard of hearing viewers' intend is to better understand the dialogues. The hypothesis is that the noise reduction and speech enhancement that Clean Audio provides will make their media experience more accessible and pleasant, allowing to perceive the dialogues easier and louder without raising, at the same time, the volume of interfering sounds. Enhanced dialogue due to Clean Audio processing is expected to improve intelligibility, thus helping the hard of hearing to better perceive the contents of the processed materials.

- Normal hearing people
  Not only people with sensory impairments can benefit from Clean Audio media products. The general audience may also find more enjoyable programmes with less noise, gratuitous ambience sounds and irrelevant background dialogues. The resulting audio mix is aimed to require less concentration, less effort, and some days that is exactly what we are looking for when relaxing in front of the screen. However, even if the general audience does not prefer Clean Audio to the current sound mix in television shows and films, it is important to consider that it should not be a disturbing factor either. It must be taken into account that many households include hard of hearing viewers and hearing audiences that may want to enjoy television together, for instance grandparents and their grandchildren. Clean Audio intends to be an inclusive access service, therefore it should allow the possibility of two people with different needs to watch the same audio visual product together and receive the same information. The aim of this test was to verify any deformation of the audio that may cause any sort of discomfort or annoyance among the general audience, in order to allow this technology to be used by hard of hearing and normal hearing users at the same time without detriment in the experience for any of them.

4.2.5. Workflow / Production Aspects

All material was prepared manually for this lab test. The selection of specific material with respect to the audio is an important issue, for not every television program is suitable for the testing. The testing material was selected from TVC's archives, from which only productions containing stereo audio were available. The following aspects must be considered to select and generate adequate testing material:

- each clip has to present both speech or dialogue and background noise or sound effects;

- this dialogue and sound effects must have been recorded separately and mixed afterwards;
- the sound effects and the dialogue ideally have to be available during the whole clip in a similar fashion;
- each clip has to be longer than 20 seconds, so that, when dividing the clip into four parts (see section 4.3.2) each part is still long enough for the test participants to get an impression of the intelligibility.

While for this sub-pilot there only was the manual workflow described in section **4.2.1**, for an actual service, the Clean Audio tool should be integrated in an existing audio workflow. This requirement has been considered during the design and development of the Clean Audio generator (see D4.2 **[2]**).

4.3.Description of User Tests

4.3.1. Aim

The lab tests included two different experiments: Clean Audio testing as described in section 4.2 and Language Learning for children testing. We will focus here on the Clean Audio testing, relating to speech intelligibility; for further details on the Language Learning test, please see section **7.3**.

- Clean Audio for hard of hearing

Aim: To gather information about whether the participants found a Clean Audio track more intelligible than the original audio mix (included in the tests as a hidden reference).
Hypothesis: Clean Audio will facilitate intelligibility for the hard of hearing, by providing a clearer and understandable audio 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) considered Clean Audio useful or, at least, inconsequential
Hypothesis: The listening experience of the hearing audience, if not improved by the increased intelligibility achieved by Clean Audio processing, will not be worsened either.

4.3.2. Methodology

The same methodology was used as in the third German lab test. This methodology was specifically designed by UAB and IRT. The overall test consisted on a set of 20 to 25-second-long video clips each divided into four parts. One part was left as the original audio visual product, and each of the remaining parts was processed in a different way with the Clean Audio tool. The clips were then shown to the users in a software environment that stops playing after each part. The users were then asked by means of a multiple-choice dialog box how good the dialogue intelligibility was for that section. Additionally, at the end of each whole clip the participants were asked to state which of the four parts they preferred. The testing software also allows for an automatic data (user feedback) gathering and later data analysis.

The methodology is described in more detail in Annex **11.2**.

4.3.3. Testers

There were two types of user groups in this test: the hard of hearing and the general audience viewers.

- The hard of hearing. 14 hard of hearing people without hearing aids were available for the tests. Each tester was presented a total of 13 clips each divided into four differently processed parts, therefore responding to the questionnaire for a total of 52 parts.

- General audience. A total of 28 normal hearing respondents participated in the lab test. Each tester was presented a total of 13 clips each divided into four differently processed parts, therefore responding to the questionnaire for a total of 52 parts.

4.3.4. Report on Test

The tests were run in Barcelona. All users were asked to go to the testing location, were the setting was the same for all of them. They were asked to sit in front of the screen and were given the information and consent forms, and then the instructions for the test were explained to them. They were asked to start the test at their earliest convenience and a member of UAB was with them to provide support if needed.

4.4.Evaluation of Sub-pilot

Both UAB and IRT participated in the evaluation of the sub-pilot. IRT collected and documented the resulting data which were then provided to UAB for further evaluation and analysis. The assessment of the collected feedback from the user tests as carried out by UAB was based on the evaluation done for the third German lab test, carried out by IRT (see section 7.1.2).

The questionnaire was broadly divided into two sections. The first section gathered personal details relevant for the analysis of the feedback such as age, gender, degree of hearing loss, and if the testers wore a hearing aid. The second section of the questionnaire was embedded in the software environment designed to collect feedback on the clips to compare the different CA processings with the original version (as a hidden reference). For further details on the testing methodology and the CA versions used in the tests, please refer to section **11.2**.

4.4.1. Results from User Feedback

RESULTS FOR THE HARD OF HEARING USERS.

All three Clean Audio versions A, B and C presented less intelligibility problems for the hard of hearing respondents than the hidden reference, see

Table 8.

Table 8. Clean Audio results for the hard of hearing.

	Hidden reference	A	B	C	Graphic comparison
I didn't understand	2	3	2	2	
It was exhausting to understand	8	8	6	6	
It was a little exhausting to understand	69	49	42	35	
I had no problem understanding	103	122	132	139	

From **Figure 10** it can be seen that, when testers considered parts a little exhausting to understand, it was more likely that such parts were the hidden reference.

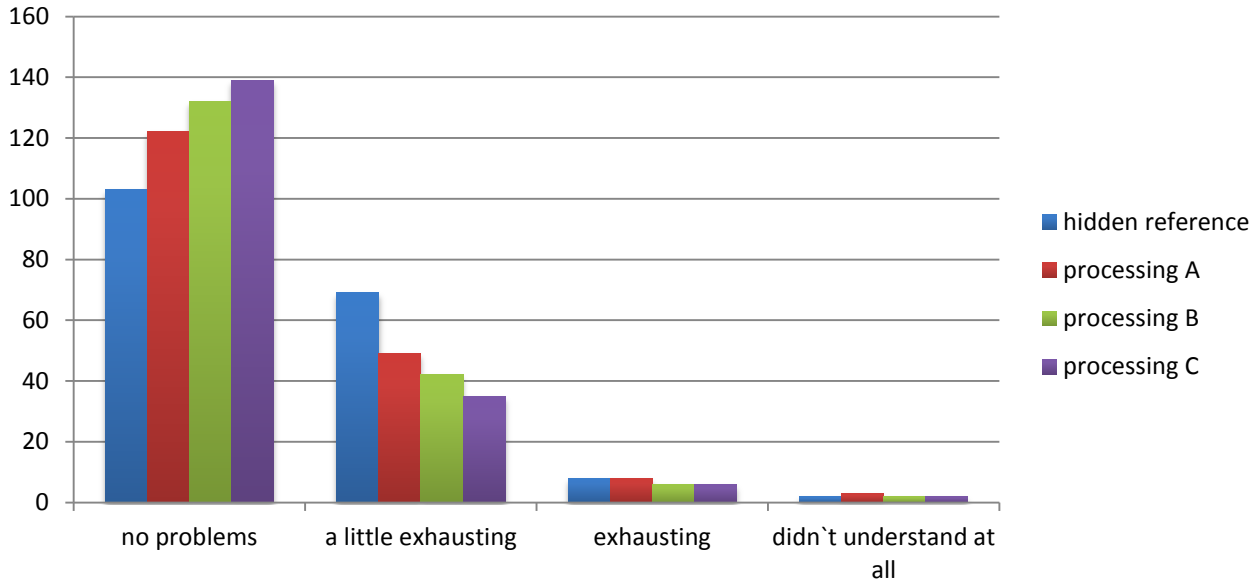


Figure 10. Clean Audio results for the hard of hearing users

When asked about which part testers preferred as far as the overall sound mix was concerned, the answers were in line with the results of the intelligibility questionnaire, C and B being the most voted processings, see **Table 9**.

Table 9. Clean Audio preference results for the hard of hearing.

	Hidden reference	A	B	C
Number of parts	32	39	51	60

However, from these figures and charts we cannot infer the improvement in speech intelligibility with the Clean Audio processing, if any. Therefore, an additional evaluation of the test results was carried out, based on a numerical rating system where 0 means: "I had no problem understanding the speech" and 3 means: "I didn`t understand the speech at all". With these numbers, the differences were calculated between the ratings given for the hidden reference and for the three different processings for each clip. For instance, when the hidden reference was rated "I didn't understand the speech at all", (3 in this numerical system), and processing A was rated "It was a little exhausting to understand the speech," (1), there would be an intelligibility improvement of 2 grading categories[5]. Based on these calculations, it was possible to show the difference in intelligibility between the original (included in the tests as "hidden reference") and the three versions of processing in each clip. To show only the clips for which an improvement in speech intelligibility could be obtained, only the differences were taken where the hidden reference was rated as "a little exhausting" or worse. I.e., only the clips where the intelligibility of the original was not excellent ("no problems" on the x-axis of the "Rating option" scale as shown in **Figure 10**) were included.

The results of this evaluation are shown in **Figure 11**, which shows that for most of the clips it was possible to improve the speech intelligibility (positive scale). Processing C indicates the best results: in 45 instances it shows improvement of intelligibility by 1 step and in 5 instances by 2 steps. Only for a few sequences the intelligibility was degraded by the CA processing (negative scale). All three processings show a lot more improvement than decrease. There are also many sequences where the intelligibility stayed the same (0), above all for processings A and B.

[5] Please note that, although a linear scale was used for the numbering system, it just identifies any improvement, but does not quantify the actual amount. E.g. an improvement of 2 is not twice better than an improvement of 1.

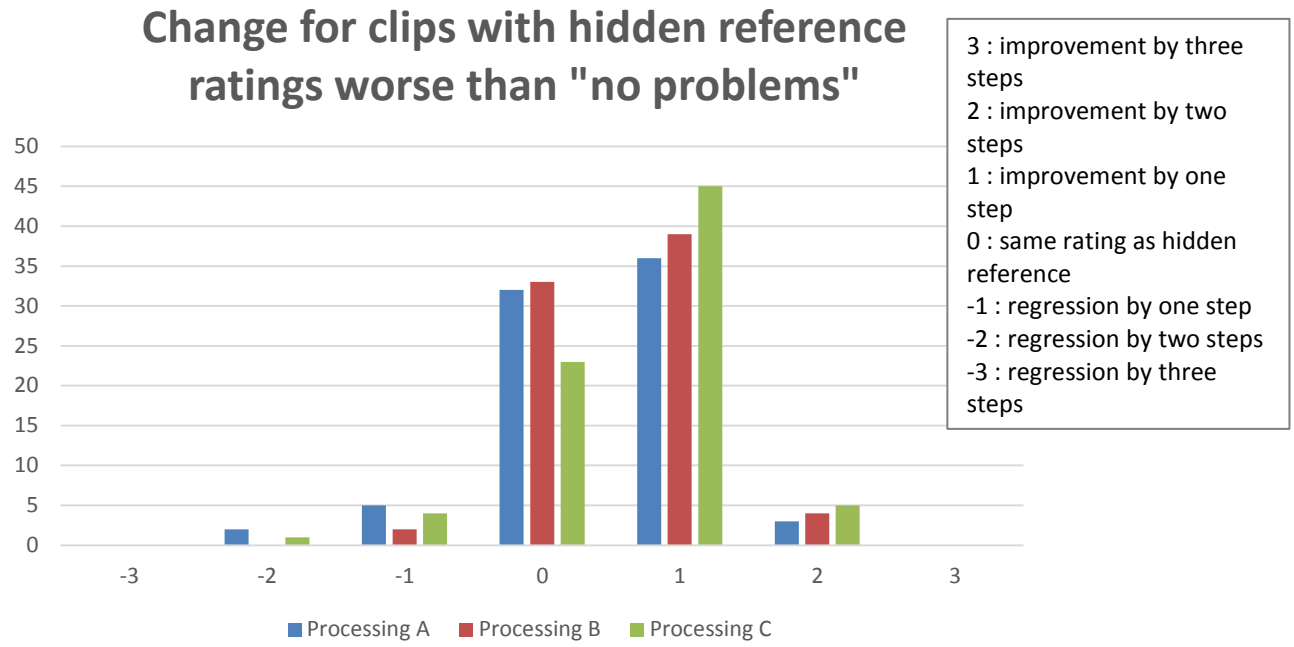


Figure 11. Clean Audio – change in intelligibility for hidden reference ratings worse than "no problems" for hard of hearing users group

RESULTS FOR THE GENERAL AUDIENCE.

Amongst the normal hearing users only on few occasions the contents of the clips were found especially hard to understand. In all cases, the original (hidden reference) showed worse scores than all three CA processings, thus implying that CA helps improving speech intelligibility also for this target group. There was no apparent difference between the three processings for this user group, see **Table 10**.

Table 10. Clean Audio results for the general audience.

	Hidden reference	A	B	C	Graphic comparison
I didn't understand	1	1	1	0	
It was exhausting to understand	21	18	15	13	
It was a little exhausting to understand	97	70	65	60	
I had no problem understanding	245	275	283	291	

In **Figure 12** it can be seen that, when testers considered the parts a little exhausting to understand, it was more likely that such parts were the hidden reference.

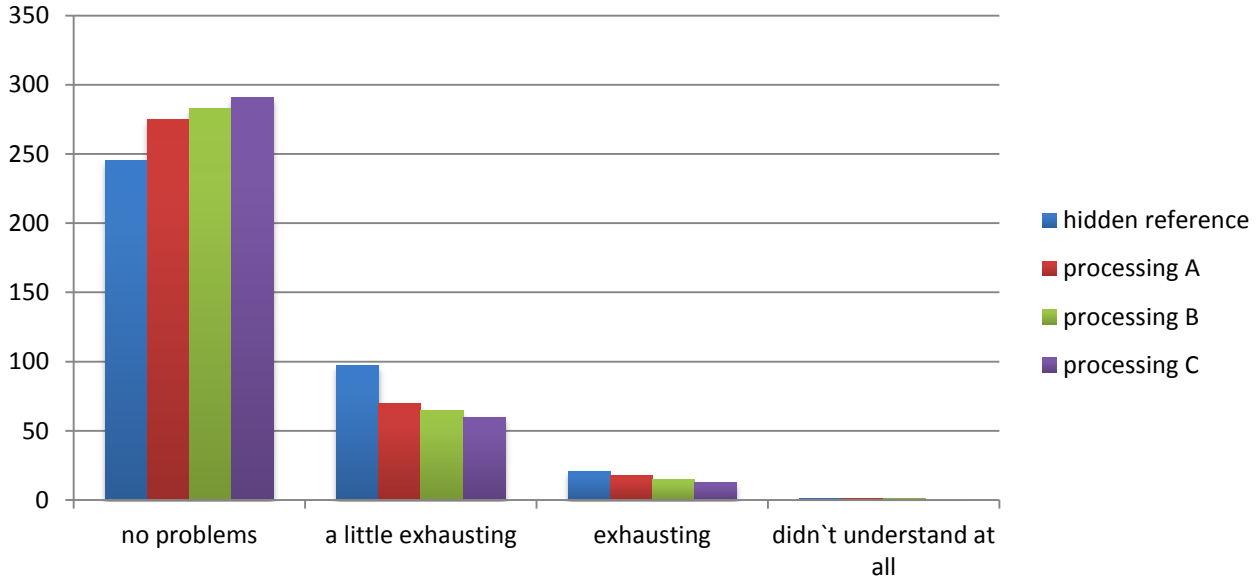


Figure 12. Clean Audio results for the normal hearing users

When asked about which part testers preferred as far as the sound mix is concerned, the results were as follows:

Table 11. Clean Audio preference results for normal hearing users

	Hidden reference	A	B	C
Number of parts	81	101	82	100

These results do not allow drawing a clear conclusion with respect to the preference of audio mix. Both CA processings A and C seem to have preference.

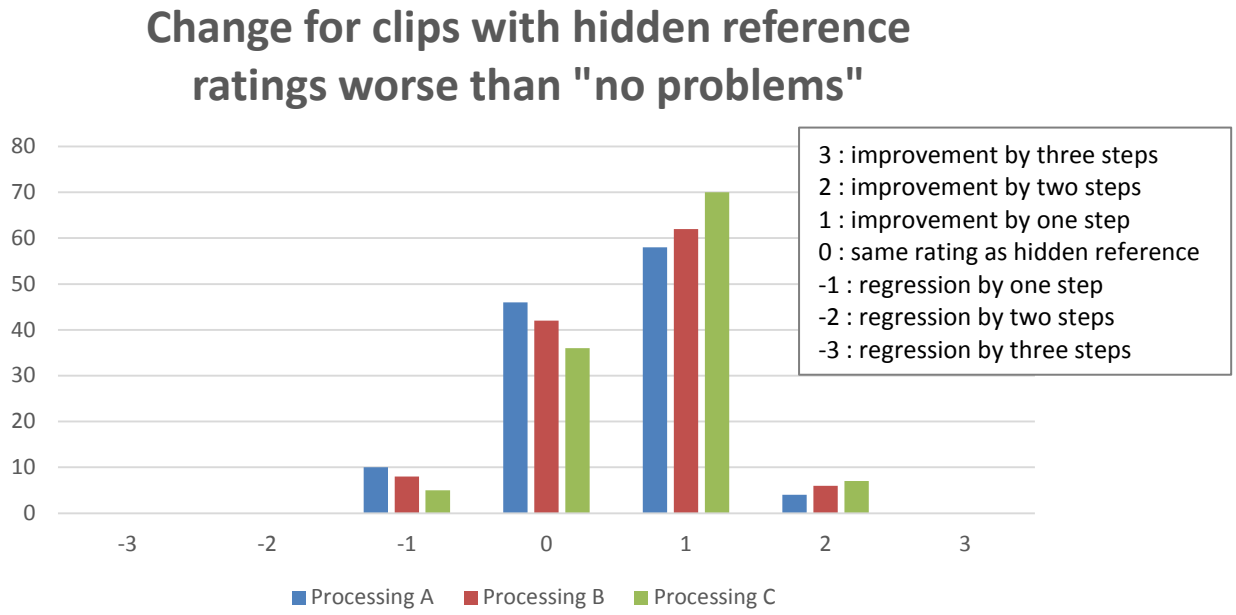


Figure 13. Clean Audio – change in intelligibility for hidden reference ratings worse than "no problems" for general audience group

Regarding the improvement of speech intelligibility, **Figure 13** indicates that by means of Clean Audio it was possible to improve the speech intelligibility (positive scale). Processing C shows the best scores, with 70 cases of improved intelligibility by 1 step and 3 cases by 2 steps. Only in a few cases the intelligibility was degraded by the CA processing (negative scale). All three processings showed more improvement than decrease. There are also many sequences where the intelligibility stayed the same (0).

4.5.Results and Insights

The results for both normal hearing and hard of hearing users indicate that applying CA improves the speech intelligibility in this lab test. All three processings indicate an improvement with the intelligibility of the audio visual materials tested. Processing C shows the best scores for both user groups. It has to be taken into account that only material containing stereo audio mix was available for the tests.

Only in very few instances CA worsened the experience, a very small number when compared to the cases in which intelligibility was improved.

One recommendation was made by one of the hard of hearing testers, who had an 80 % hearing loss and was wearing a hearing aid. This user suggested to combine the Clean Audio service with the already available subtitle service, which from a technical point of view is indeed no problem whatsoever. According to this user, the combination of both will be of great help for people with her condition. There are people who can still hear something and want to take advantage of this possibility, but whose hearing loss still requires for them to read subtitles in order not to miss any dialogue contents. Clean Audio, this user claimed, could be a good complement to their user experience, since they could hear more of the dialogue and still rely on the subtitles in case they miss out on some parts.

5. Audio Description Sub-Pilot in Spain (Catalonia)

During 2015 and the first quarter of 2016 TVC's Audio Description service was extended from traditional broadcast TV to HbbTV services. This process required several changes in the TVC production workflow, particularly in the data structure, coding and publishing of assets. Once these changes were completed, the new service was opened to the public, allowing this sub-pilot to run from April to September 2016.

5.1. Goals of Sub-pilot

This sub-pilot has helped to expand TVC's workflow to support provision of multiple audio tracks in HbbTV services, including Audio Description. The control of audio used for each video played has allowed direct feedback about usage of AD service from final users.

5.2. Description of service / application

5.2.1. Technical implementation

TVC developed a broadband access content delivery system based on their workflow for accessibility service offers in broadcast services. The new system allows the export of Audio Description content and merges it with original program audio to generate a new VoD asset. More information about technical implementation is explained in Annex **0** and D4.2 **[2]**.

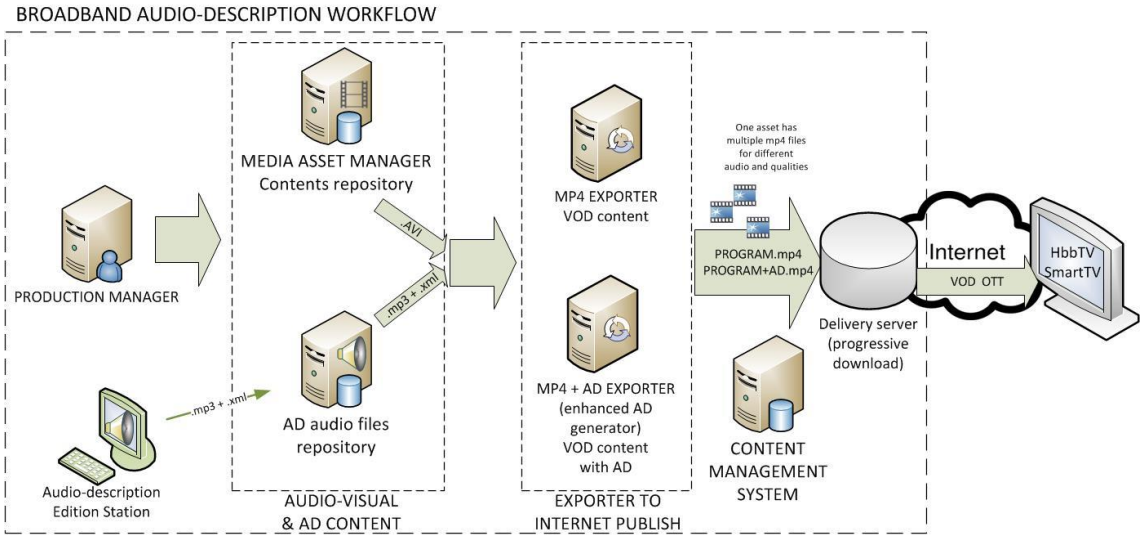


Figure 14. Basic scheme of AD workflow

The new solution publishes two assets:

- One VoD asset with original program video and dubbed Catalan audio track.
- One VoD asset with original program video and a new audio track in which Audio Description sentences are mixed with the Catalan audio.

Users access to “TV3alaCarta” HbbTV service and choose the media they want to replay. Once the media is loaded at the player, the available audios for this content are offered. Then users can choose what audio want to play.

5.2.2. Functionalities

TVC added Audio Description (and Original Sound Track, see chapter **6**) accessibility features to its HbbTV media player.

Figure 15. Audio Description option available on HbbTV media player

With this new development, viewers can find a button on the media player interface that informs about the availability of Audio Description content; in case Audio Description is not available, the button does not appear, see **Figure 15**. The Audio Description button will show the current state configuration, as shown in **Figure 16**.

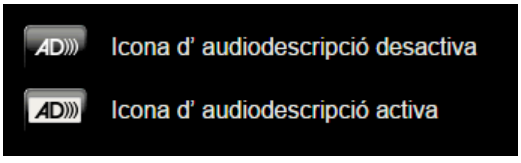


Figure 16. Audio Description de/active button states

This feature allows viewers to enable or disable Audio Description, and this choice is saved as a cookie for future access to new programs with AD contents.

The Original Sound Track (OST) feature in the HbbTV application works similarly, as explained in section **6.2.2**. However, as both audio contents cannot be played back together, TVC made an implementation to have the AD feature prevailing over OST. If the user deactivates AD, then default audio or OST will replay depending on the last saved preference.

5.2.3. Availability of Service

The Audio Description offer for TVC's Catch-Up service on HbbTV has been publicly available from April 2016 onwards and will remain operational indefinitely. Currently the TVC accessibility department provides Audio Description for broadcast in the daily own produced soap opera "La Riera" and drama weekly series "Cites", in addition to the Friday night externally produced prime time movie. For broadband, currently, only TVC's own produced daily soap opera and drama weekly series are accessible with AD, because TVC does not have the rights to offer AD on external productions available as part of its Catch-Up service.

5.2.4. Intended Audience

Whereas this access service is addressed to viewers with visual impairments, the AD functionality is available open to everybody.

5.2.5. Workflow / Production Aspects

TVC's Product Management System was initially designed for broadcast. So it was necessary to implement new options and tools to allow broadband contents programming.

TVC developed a new module in the workflow, the Multiple Audio Asset Generator (as described in **D4.2**, section **2.3**), which manages all the AD files, merges with the original audio and creates new audio track.

With new developments and changes implemented on the system it is now possible to program broadband and broadcast access contents independently, considering that access content rights may be different for broadcast and broadband distribution.

Other modules in the access workflow, like MAM (Media Asset Management), export transcoders or CMS (Content Management Systems) were conveniently modified and improved to support and publish access contents to "TV3alaCarta" VoD HbbTV service.

5.3.Description of Sub-Pilot Tests

5.3.1. Aim

In order to collect quantitative data on usage of service, a newly developed JavaScript code runs from within TVC's HbbTV VoD application and gathers all events like pressed buttons, played videos, etc. This gathering is done anonymously on an Adobe Omniture service (as described in D4.1, section **4.3.5**) **[1]**, so it does not gather any personal information but only quantitative data that will help to measure and analyse the audience.

5.3.2. Methodology

The evaluation was conducted by measuring and analysing the activity of the audience. These measurements were performed using JavaScript to send viewing events to Adobe Omniture platform, used for audience measurement purposes (as described in D4.1 **[1]**, section **4.3.5**). This mechanism allows collecting information on the number of play backs and hours of video consumed for each specific content, which includes the type of audio that was played back. Specifically the audio chosen for each video played is stored in an "Audio" variable that can have the values "STD" for Catalan audio track, "AUD" for Audio

Description track or "VOR" for Original Sound Track. Moreover, information on unique users who use the application can be extracted, but this data cannot be completely accurate as it depends on the TV device configuration that must accept navigation cookies. When cookies are not accepted then it's not possible consolidate the interactions done by the device and every connecting session is counted as a new unique device.

The pilot had a foreseen duration of six months but finally only four months of data have been correctly gathered. In mid-August a change in TVC's Content Management System was applied. This change caused the HbbTV service to stop collecting usage data. This problem was detected mid-September, then it was solved resuming the gathering data use. Because of this, the quantitative data analysed only covers the months April 2016 till July 2016.

5.3.3. Testers

Any user of "TV3alacarta" Catch-Up can access the service and consequently was a potential tester. The only limitations are due to the geographical broadcast signalization (which also signals the HbbTV application) only available in Catalonia. Currently, AD is only offered over broadcast and HbbTV broadband channels. However, the AD implementation on web Catch-Up service is expected shortly, thanks to new developments and changes implemented on access contents workflow.

5.3.4. Report on Test

As detailed in section **5.3.2**, due to changes applied at the TVC HbbTV service in mid-August 2016, the mark-up data about the use of audio services for the month of August and September are incomplete or non-existent. Therefore, the correctness of this data is not guaranteed, and it was decided not to use it in the evaluation. Finally, the evaluation has been extended to include the information gathered during the months from April to November, discarding months of August and September. Moreover, the data from November are only for the first third of the month and were extrapolated to cover the entire month of November.

5.4.Evaluation of Sub-Pilot

There was a clear limitation implied by the Sub-Pilot period: the seasonal usage caused decrease in usage because of the summer holiday. The summer 2016, included in this trial, shows a descend in all parameters monitored in this test. The seasonal effect due to summer season is shown for last year (2015) and present year (2016) in **Figure 17**.

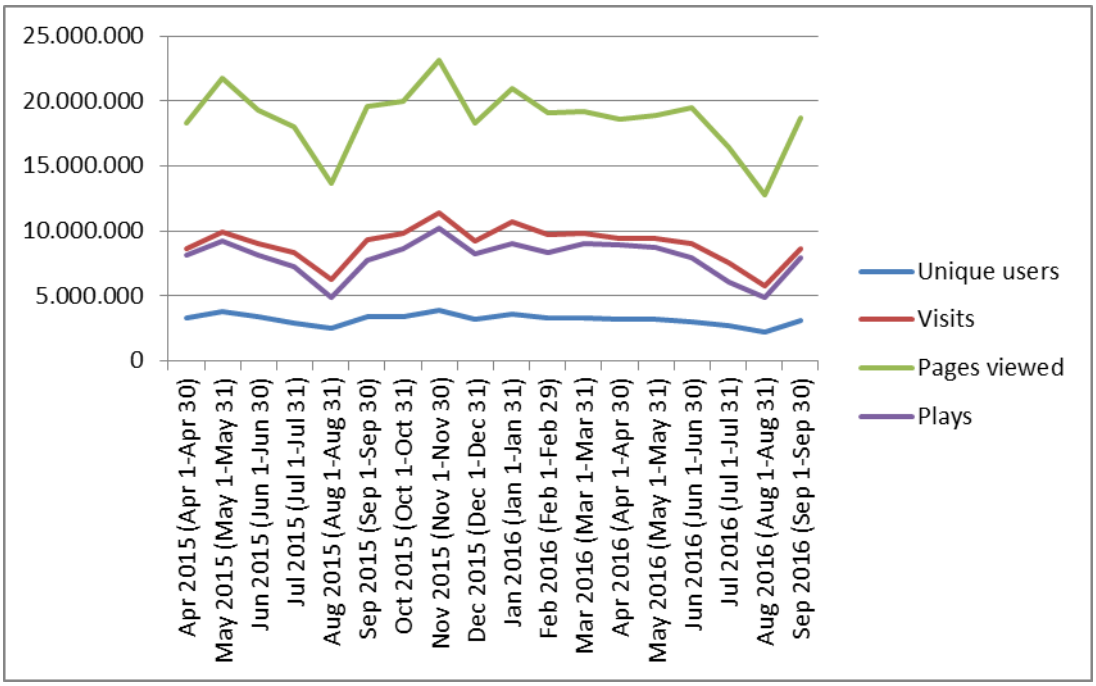


Figure 17. TVC IP total usage data

The chart analysis shows that IP consumption in general, for any of the parameters, decreases in June, July and August due to the summer holidays. It can also be observed that this decline is close to 25% of pages viewed and around 30-40% in visits and plays.

Moreover, it should be considered that the AD service is offered only three products in TVCs weekly program:

- Daily afternoon soap opera,
- Prime time soap opera on Mondays
- Prime time movie on Fridays.

The third product cannot be found to TVC Catch-Up TV service, due to rights issues. The other two products interrupt their normal cycle from the last week of June until the end of August. Thus, during these two months the products available with AD had already been available via broadcast and broadband previously, in some cases with exactly the same format. This may impact the number of requests, as it can be expected that viewers may not be interested in seeing the same product again.

Figure 18 shows the detailed availability of these AD products in the Catch-Up service along the sub-pilot period:

- The daily afternoon soap opera ("La Riera – Setena temporada", in color green) transforms to a best-moments product with the most significant scenes from the last year, it's like a pseudo-reposition.
- Finally, the weekly soap opera ("Cites" and "Nit i dia", in colours yellow and blue respectively) is substituted by repositions from other older soap opera products with the same format.

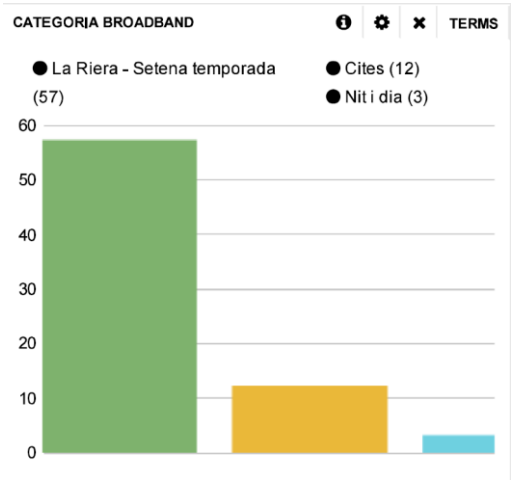


Figure 18. Amount of AD programs available in broadband, for months April to July

5.4.1. Quantitative evaluation

As a resume we have prepared data gathered in

Table 12 and Table 13 with the most significant values for the period comprised by the sub-pilot. For the average calculation purposes the months of August and September have been rejected.

"Plays" are the total reproductions made, regardless of the length of view, "Plays AD" meaning the playback was with AD track activated and "Plays HbbTV" referring to all reproductions made, irrespective of the audio option chosen.

The "Hours viewed" are the sum of hours of the whole time in which the videos have been played, where "Hours viewed AD" means videos were played with AD audio track and "Hours viewed HbbTV" refers to all videos viewed irrespective of the audio playback chosen.

Table 12. Usage of Audio Description during pilot period compared with total HbbTV usage in Plays and % Hours viewed

	April	May	June	July	August	September	October	November	Average
Plays AD	2.726	2.251	1.228	614	156	130	859	1.761	1.573
Plays HbbTV	444.573	397.699	355.509	243.345	176.808	351.922	422.745	409.437	378.885
% Plays AD vs HbbTV	0,61%	0,57%	0,35%	0,25%	0,09%	0,04%	0,20%	0,43%	**0,40%**
Plays HbbTV AD available	217.117	205.102	189.495	79.774	19.906	149.360	196.774	227.949	186.035
% Plays AD vs HbbTV AD available	1,26%	1,10%	0,65%	0,77%	0,78%	0,09%	0,44%	0,77%	**0,83%**
% Plays HbbTV AD available vs Plays HbbTV	48,84%	51,57%	53,30%	32,78%	11,26%	42,44%	46,55%	55,67%	**48,12%**
Hours viewed AD	367	389	195	99	25	38	106	205	227
Hours viewed HbbTV	114.050	100.490	101.412	70.752	519.877	102.297	125.970	140.751	108.904
% Hours viewed AD vs HbbTV	0,32%	0,39%	0,19%	0,14%	0,00%	0,04%	0,08%	0,15%	**0,21%**
Hours vieved HbbTV AD available	54.892	56.081	57.533	26.755	7.181	51.764	26.755	90.837	52.142
% Hours viewed AD vs HbbTV AD available	0,67%	0,69%	0,34%	0,37%	0,35%	0,07%	0,40%	0,23%	0,45%
% Hours viewed HbbTV AD available vs Hours viewed HbbTV	48,13%	55,81%	56,73%	37,82%	1,38%	50,60%	21,24%	64,54%	**47,38%**

"Unique TVs" refers to identifiers sent to the TV as cookies, if the configuration of the device allows it: this cookie is preserved over time, uniquely identifying the TV device. In this case "Unique TV AD" means devices that had played videos with AD and "Unique TV HbbTV" includes all devices which have played any video, AD or not.

Finally, "Visits" are the total times a page in TVCs Catch-Up service was visited, regardless if the media on that page was played and of the duration of the reproduction.

Table 13. Usage of Audio Description during pilot period compared with total HbbTV usage in Unique TVs and Visits

	April	May	June	July	August	September	October	November	Average
Unique TVs AD	656	541	364	270	65	44	398	654	481
Unique TVs HbbTV	48.837	46.130	44.477	37.016	26.333	49.792	57.568	97.428	55.243
% Unique TVs AD vs HbbTV	1,34%	1,17%	0,82%	0,73%	0,25%	0,09%	0,69%	0,67%	**0,90%**
Unique TVs HbbTV AD available	22.353	24.059	24.311	15.014	4.734	24.778	31.468	63.735	30.157
% Unique TVs AD vs HbbTV AD available	2,93%	2,25%	1,50%	1,80%	1,37%	0,18%	1,26%	1,03%	1,79%
% Unique TVs HbbTV AD available vs Unique TVs HbbTV	45,77%	52,15%	54,66%	40,56%	17,98%	49,76%	54,66%	65,42%	**52,20%**
Visits AD	1.272	1.088	647	335	91	86	486	903	789
Visits HbbTV	190.208	171.535	161.996	108.728	75.380	169.264	206.367	222.999	176.972
% Visits AD vs HbbTV	0,67%	0,63%	0,40%	0,31%	0,12%	0,05%	0,24%	0,40%	0,44%
Visits HbbTV AD available	92.766	91.660	88.163	36.265	9.268	81.465	109.939	136.533	92.554
% Visits AD vs HbbTV AD available	1,37%	1,19%	0,73%	0,92%	0,98%	0,11%	0,44%	0,66%	0,89%
% Visits HbbTV AD available vs Visits HbbTV	48,77%	53,44%	54,42%	33,35%	12,30%	48,13%	53,27%	61,23%	50,75%

From

Table 12 and Table 13 the following can be seen: taking with care due the cookies acceptance influence explained at section 5.3.2, approximately half **(**52,20%) of TV devices have been in some moment watching some media offering the possibility of AD. From the total plays done, 48,12% played medias offering the AD option. Finally, from all hours views the AD were available in the 47,38% of the medias viewed. From the total HbbTV devices that receive TVC's HbbTV service, 0,90% of the TV devices have played in some moment of the month a product with AD active. 0,40% of the total plays done in HbbTV were with AD activated, that are 0,21% of the total hours viewed in TVC's HbbTV service.

Based on these numbers TVC already considers a significant AD coverage, being about half of total watching preferences of the users.

5.5.Results and Insights

From the analysis of the data an acceptance and regular use of the TVC AD Catch-Up service can be observed, for a clear and stable group of users. Therefore, TVCs internal recommendation is to maintain the AD service in TV3AlaCarta and to expand the products (assets) offered when possible.

It's important remark that the 0,44% of Catalonia populations has some kind of recognized visual disability:

33.091	Visual dissability recognized, year 2015 (http://www.idescat.cat/pub/?id=aec&n=847&lang=cat)
7.508.106	Catalunya population, year 2015 (http://www.idescat.cat/pub/?id=aec&n=245)
0,44%	**Visual dissability respect total population in Catalonia, year 2015**

Although there is no data available about the usage of AD in broadcast, from this premises it seems feasible to have a usage of AD in Catch-Up TV rounding to **0,80%** when it is available **(0,83% Plays AD vs HbbTV AD available in average)** and not more penetration can be expected.

6. Other Languages Sub-pilot in Spain (Catalonia)

During 2015 and the first quarter of 2016 TVC original language service was extended from traditional broadcast TV to HbbTV services. This process required several changes in the TVC production workflow, particularly in the data structure, coding and publishing of assets. Once these changes were completed, the new service was opened to the public, allowing this sub-pilot to run from April to September 2016.

6.1. Goals of Sub-Pilot

This audio Sub Pilot has helped to expand the workflow system to support multiple audio on HbbTV services (including the Original Sound Track, OST). Markup implementation has allowed direct feedback about acceptance and usability of OST service from final users.

6.2. Description of service / application

6.2.1. Technical implementation

TVC expanded the contents export system to generate new VoD assets with OST, see **Figure 19**.

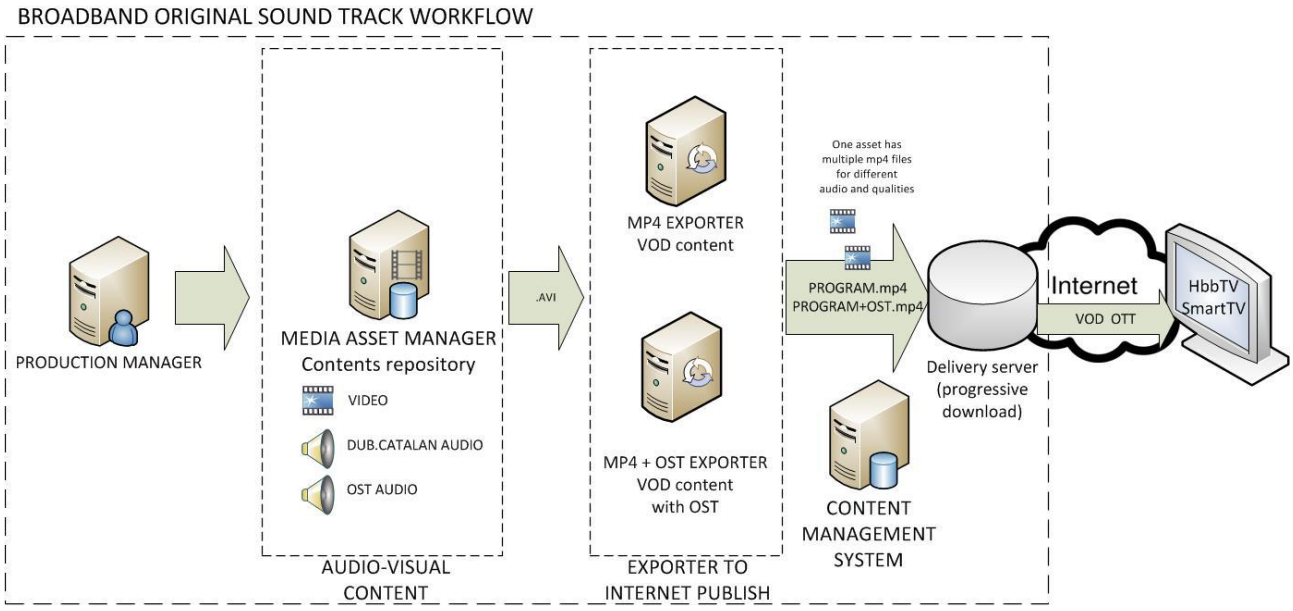


Figure 19. Basic plan OST workflow

The new solution publishes two assets:

- One VoD asset with original program video and dubbed Catalan audio track.
- One VoD asset with original program video and the OST (for example, English soundtrack for an English film).

Users access to "TV3alaCarta" HbbTV service and choose the content they want to replay.

6.2.2. Functionalities

TVC added Audio Description and Original Sound Track accessibility features to its HbbTV media player.

Figure 20. Original Sound Track option available on HbbTV media player

With this new development, viewers can find a button on the media player interface that informs about the availability of OST content; in case Original Sound Track is not available, the button does not appear, see **Figure 20**. The OST button will show the current state configuration, as shown in **Figure 21**.

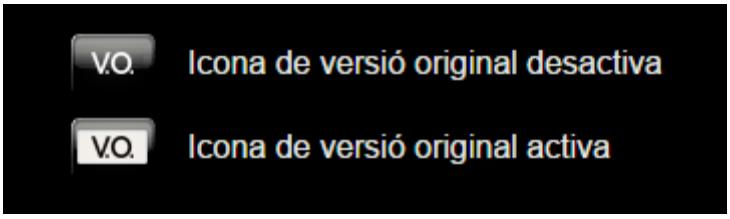


Figure 21. Original Sound Track de/active button states

This feature allows viewers to enable or disable OST, and this choice is saved as a cookie for future access to new programs with OST contents.

The Audio Description feature works similar, as explained in section **5.2.2**. However, as both audio contents cannot be played back together, the AD feature prevails over OST. If the user deactivates AD, then default audio or OST will replay depending on the last saved preference.

6.2.3. Availability of Service

The Original Sound Track offer for TVC's Catch-Up service on HbbTV has been publicly available from April 2016 onwards and will remain operational indefinitely. Currently the TVC accessibility department provides OST for broadcast in the daily externally produced soap opera "El Faro", kids serial "Bubble Bip" in addition to some externally produced prime time movies. Currently, only "El Faro" and "Bubble Bip" contents are accessible in TVC VoD service with OST, because TVC does not have the rights to make OST on other productions available as part of the Catch-Up service.

6.2.4. Intended Audience

This service is open to everybody and is specially addressed to people who want to hear OST for language learning purposes or for non-Catalan speakers who want to access TVC contents.

6.2.5. Workflow / Production Aspects

TVC's Product Management System was initially designed for broadcast. So it was necessary to implement new options and tools to allow broadband contents programming.

With new developments and changes implemented in the system it is now possible to program broadband and broadcast access contents independently, considering that access content rights may be different for broadcast and broadband distribution.

Other modules in the access workflow, like MAM (Media Asset Management), export transcoders or CMS (Content Management Systems) were conveniently modified and improved to support and publish access contents to "TV3alaCarta" VoD HbbTV service.

6.3.Description of Sub-pilot Tests

6.3.1. Aim

In order to collect quantitative data on usage of service, a newly developed JavaScript code runs from within TVC's HbbTV VoD application and gathers all events like pressed buttons, played videos, etc.

This gathering is done anonymously on Adobe Omniture service (as described in D4.1 **[1]**, section **4.3.5)**, so it does not gather any personal information but only quantitative data that will help to measure and analyse the audience.

6.3.2. Methodology

Whereas initially a six months' period was planned for this sub-pilot, finally for four months, from April 2016 till July 2016 (see section **6.3.4**), quantitative data was gathered on an Adobe Omniture allowing a precise analysis of service impact on final users (as described in D4.1 **[1]**, section **4.3.5**).

6.3.3. Testers

OST has been deployed as part of the public TVC HbbTV VoD service as of April 2016. This allowed testing in a real public open service, open to all public.

For testing purposes it is important to consider next points:

- The OST service was not specifically advertised, so the users only discovered the OST option through the new OST button when playing a media for which OST audio content was available.
- Currently, only few TVC VoD contents have OST audio, due to rights' issues with these contents.

6.3.4. Report on Test

As explained in section **5.3.2**, due to changes applied in mid-August at the TVC HbbTV service, the mark-up data about the use of audio services for the month of August and September are inconsistent and could not be used. Therefore, the evaluation could only include the information gathered during the months from April to July 2016. To compensate for this, the data collected during the months of October and first third of November have been added to the study. The data of November have been extrapolated to cover the whole month.

6.4.Evaluation of Sub-pilot

In the same way as detailed in section **5.4**, the holiday period affected the results obtained during the pilot. Moreover, the case of Original Sound Track has the limitation concerning to rights' issues and reduced availability of products for this service: TVC during the sub-pilot only had one product available. It was a soap opera "El faro, cruilla de camins" with the Galician Original Sound Track.

6.4.1. Quantitative evaluation

The following **Table 14** and **Table 15** presents a summary of the results obtained, later in the text are reviewed data relating to the average values obtained during this extended pilot.

"Plays" are the total reproductions made, regardless of the length of view, "Plays OST" meaning the playback was with OST track activated and "Plays HbbTV" referring to all reproductions made, irrespective of the audio option chosen.

The "Hours viewed" are the sum of hours of the whole time in which the videos have been played, where "Hours viewed OST" means videos were played with OST audio track and "Hours viewed HbbTV" refers to all videos viewed irrespective of the audio playback chosen.

Table 14. Statistics of Original Sound Track in Plays and Hours viewed

	April	May	June	July	August	September	October	November	Average
Plays OST	32	75	197	68	0	0	19	78	78
Plays HbbTV	444.573	397.699	355.509	243.345	176.808	351.922	422.745	409.437	378.885
% Plays OST vs HbbTV	0,01%	0,02%	0,06%	0,03%	0,00%	0,00%	0,00%	0,02%	**0,02%**
Plays HbbTV OST available	11.166	10.520	10.119	5.942	297	409	8.382	15.426	10.259
% Plays OST vs HbbTV	0,29%	0,71%	1,95%	1,14%	0,00%	0,00%	0,23%	0,51%	0,80%
% Plays HbbTV OST available vs Plays HbbTV	2,51%	2,65%	2,85%	2,44%	0,17%	0,12%	1,98%	3,77%	**2,70%**
Hours viewed OST	3	11	49	21	0	0	5	18	18
Hours viewed HbbTV	114.050	100.490	101.412	70.752	519.877	102.297	125.970	140.751	108.904
% Hours viewed OST vs HbbTV	0,003%	0,01%	0,05%	0,03%	0,00%	0,00%	0,00%	0,01%	0,02%
Hours vieved HbbTV AD available	4.672	5.224	5.546	3.008	0	0	3.375	5.517	4.557
% Hours viewed AD vs HbbTV AD available	0,06%	0,21%	0,88%	0,70%	0,00%	0,00%	0,15%	0,33%	0,39%
% Hours viewed HbbTV AD available vs Hours viewed HbbTV	4,10%	5,20%	5,47%	4,25%	0,00%	0,00%	2,68%	3,92%	**4,27%**

"Unique TVs" refers to identifiers sent to the TV as cookies, if the configuration of the device allows it, this cookie is preserved over time, uniquely identifying the TV device. In this case "Unique TV OST" means devices that had played videos with OST and "Unique TV HbbTV" includes all devices which have played any video, OST or not.

Finally, "Visits" are the total times a page in TVCs Catch-Up service was visited, regardless if the media on that page was played and of the duration of the reproduction.

Table 15. Statistics of Original Sound Track in Unique TVs and Visits

	April	May	June	July	August	September	October	November	Average
Unique TVs OST	22	45	77	38	0	0	13	69	44
Unique TVs HbbTV	48.837	46.130	44.477	37.016	26.333	49.792	57.568	97.428	55.243
% Unique TVs OST vs HbbTV	0,05%	0,10%	0,17%	0,10%	0,00%	0,00%	0,02%	0,07%	**0,09%**
Unique TVs HbbTV AD available	1.924	2.082	1.767	1.350	0	0	3.368	6.264	2.793
% Unique TVs AD vs HbbTV AD available	1,14%	2,16%	4,36%	2,81%	0,00%	0,00%	0,39%	1,10%	1,99%
% Unique TVs HbbTV AD available vs Unique TVs HbbTV	3,94%	4,51%	3,97%	3,65%	0,00%	0,00%	5,85%	6,43%	**4,73%**
Visits OST	26	59	109	47	0	0	13	69	54
Visits HbbTV	190.208	171.535	161.996	108.728	75.380	169.264	206.367	222.999	176.972
% Visits OST vs HbbTV	0,01%	0,03%	0,07%	0,04%	0,00%	0,00%	0,01%	0,03%	0,03%
Visits HbbTV AD available	6.950	7.364	7.271	4.129	0	0	5.180	8.376	6.545
% Visits AD vs HbbTV AD available	0,37%	0,80%	1,50%	1,14%	0,00%	0,00%	0,25%	0,82%	0,81%
% Visits HbbTV AD available vs Visits HbbTV	3,65%	4,29%	4,49%	3,80%	0,00%	0,00%	2,51%	3,76%	3,75%

From **Table 14** and **Table 15** the following can be seen: 4,73% of TV devices have been in some moment watching some media offering the possibility of OST. Looking the total plays done, the 2,70% have played medias offering the OST option. Finally, from all hours views the OST were available in the 4,27% of the medias viewed.

From the total HbbTV devices that receive TVC's HbbTV service, 0,09% of the TV devices have played in some moment of the month a product with OST activated. 0,02% of the total plays done in HbbTV were with OST activated; this number coincides with the percentage of hours viewed with OST in TVC's HbbTV service.

6.5. Results and insights

With this integration the target of improving the accessibility by offering more languages have been achieved. It is therefore highly recommended the maintenance of the service beyond the pilot project HBB4ALL.

In order to successfully exploit the service, it is necessary to find a technical implementation that ensures the geographical broadcast signalization is really only available in Catalonia. With such a solution, the content's rights limitation disappears and more contents in OST could be offered without the current restrictions.

7. Complimentary User Tests

7.1. Clean Audio: Third and Final Lab Test

To evaluate the potential benefit of the Clean Audio generator for stereo signals, additional listening tests with both hearing impaired users with and without hearing aid were conducted. The outcome of these final lab tests is presented here. The results for 5.1 content had already shown 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**[1]**).

7.1.1. Rationale

During the second series of lab tests in March 2015, focussing on stereo content, the chosen methodology and test items did not allow to conclusively derive generic and stable knowledge regarding the effect of the Clean Audio processing. Also, some of the test items were reported as "not being critical" (meaning that in their original version they already had a good speech intelligibility – for such clips no improvement can be reached by the CA generator). It was therefore concluded that that additional, unforeseen tests had 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 (see D4.2, chapter 5.1.1 **[2]**)

These third lab tests were conducted in October/November 2015. First a large set of critical test items was identified by an intensive search in media archives (in close cooperation with RBB). Additionally, in close cooperation with UAB, a new testing methodology was developed, to eliminate the weak points of the testing methodology that had been used up till then. The methodology is described in detail in Annex **11.2**.

7.1.2. Test results

In the lab tests, nine test subjects participated with a hearing aid, and 22 without a hearing aid.

Participants without hearing aid

The test clips were rated qualitatively, with four rating options: "I had **no problem** understanding the speech", "It was **a little exhausting** for me to understand the speech", "It was **exhausting** for me to understand the speech" and "I **didn`t understand the speech at all**". First, all given ratings for all test clips were summed up for each of the four tested versions (original audio version included as "hidden reference" and three different processing; for technical details to the processed versions please refer to the description of the testing methodology in Annex **11.2** resulting in **Figure 22**. There it can be seen that for all four versions, including the hidden reference, most given ratings are "no problem", meaning that most of the used clips were already intelligible even without the processing. The processing B has most ratings at "no problem" and the fewest scores for the other three rating options.

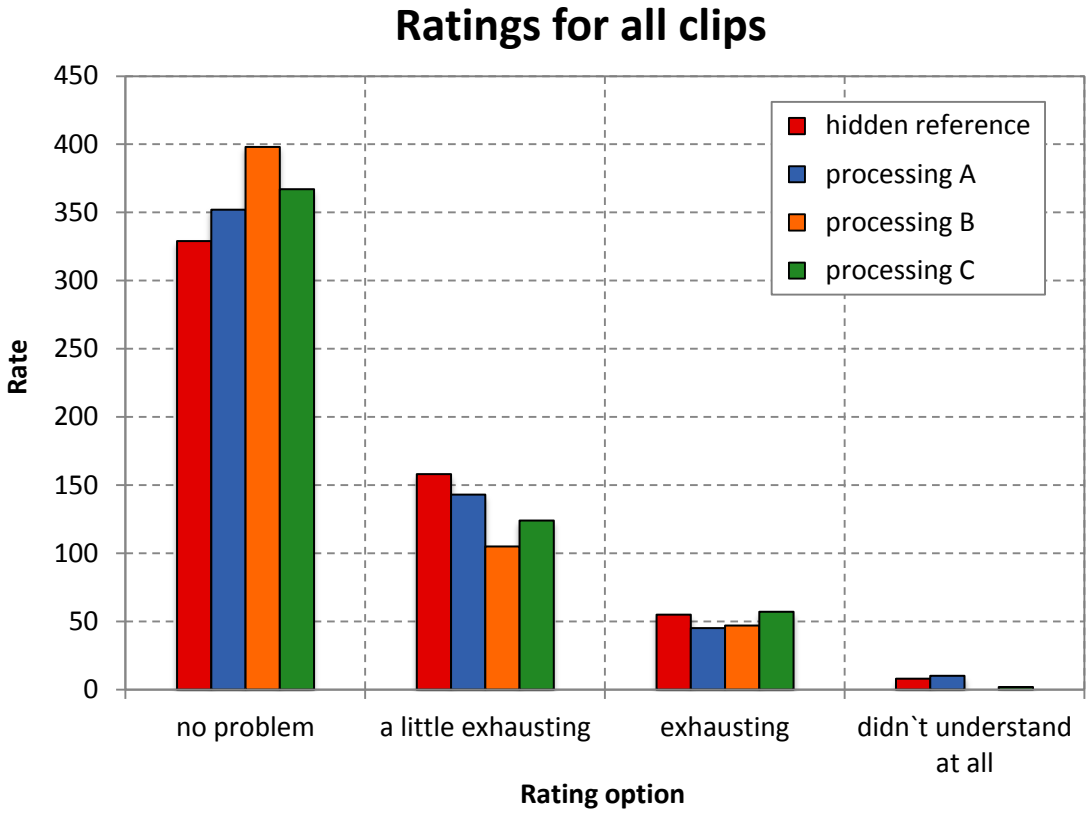


Figure 22. Clean Audio final lab test – ratings for all clips

Since the graphical evaluation doesn't conclusively show if one of the four versions was better than the other or even if there was any considerable difference between the versions, several statistical analyses were carried out with the gained data. The analysis showed that there was a significant difference between the hidden reference and the processing B as well as the processing C. Between the hidden reference and the processing A there was no significant difference. There also was none between processing B and C.

Whereas this analysis shows statistical differences between the versions, it does not tell us if it was possible to gain an improvement in speech intelligibility with the Clean Audio processing. Therefore, an additional evaluation of the test results was carried out. In this second step each of the four rating options was given a number from 0 to 3, 0 meaning: "I had no problem understanding the speech" and 3 meaning: "I didn`t understand the speech at all". With these numbers the differences were calculated between the ratings given for the hidden reference and for the three different processing for each test sequence. As a result, it was possible to show the difference in intelligibility between the original (included in the tests as "hidden reference") and the three versions of processing in each test sequence. To show only the sequences for which an improvement in speech intelligibility could be obtained, only the differences were taken where the hidden reference was rated as "a little exhausting" or worse. I.e., only the sequences where the intelligibility of the original was not excellent ("no problem" on the x-axis of the "Rating option" scale as shown in **Figure 22**) were included.

The results of this evaluation are shown in **Figure 23**, which shows that for most of the sequences it was possible to improve the speech intelligibility (positive scale). All three processing nearly show equally good results. Only for a few sequences the intelligibility was degraded by the CA processing (negative scale).

There are also many sequences where the intelligibility stayed the same (0). Especially processing B and C show a lot more improvement than decrease (-1) or the same intelligibility (0). Statistical analysis on these numbers showed, that in this evaluation the results for processing B and C were significantly different from the results for the processing A, but not significantly different from each other.

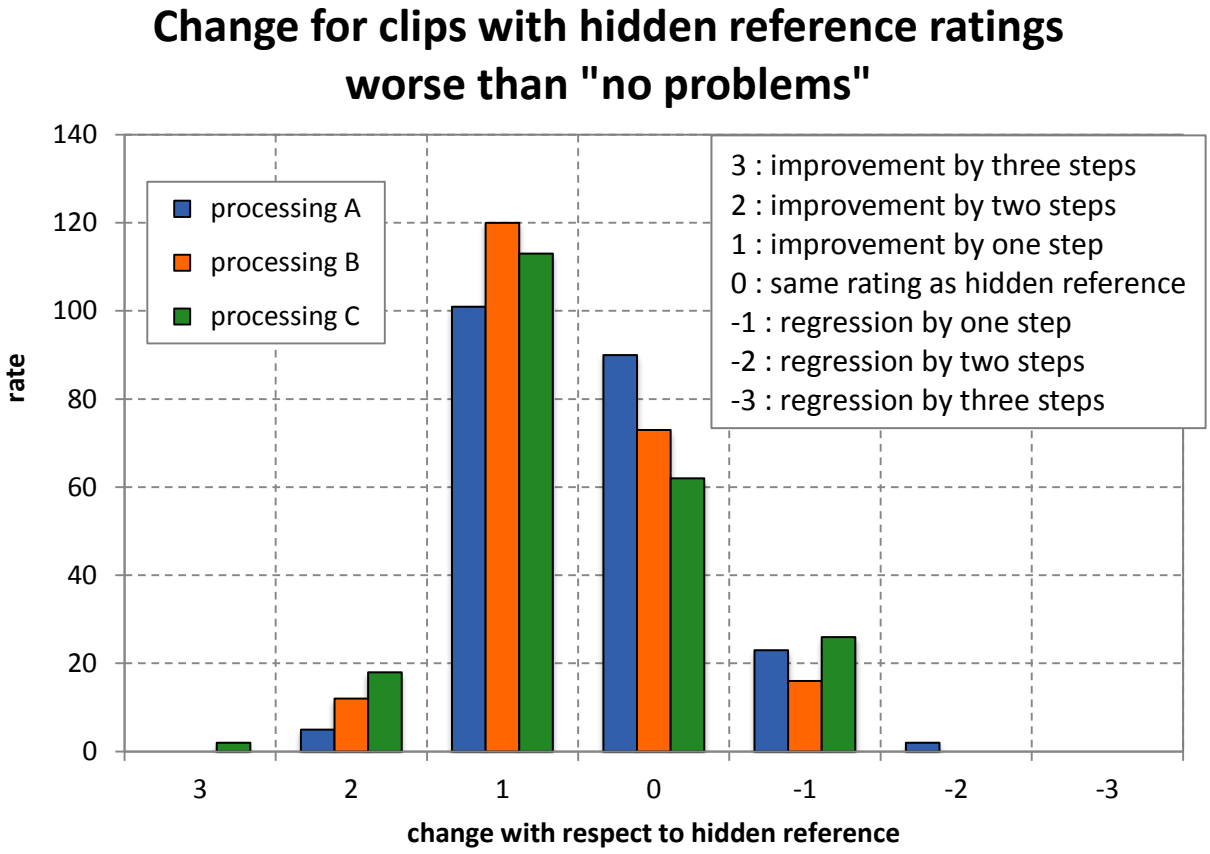


Figure 23. Clean Audio final lab test – change in intelligibility for hidden reference ratings worse than "no problems"

Participants with hearing aid

The results from this lab test for people with hearing aid are not significant because there were not enough participants in this group and the results varied heavily. This is also caused by the large variety between hearing impairments, hearing aid devices and their modes of operation.

7.1.3. Recommendations

The lab tests showed an improvement in speech intelligibility amongst the hearing impaired participants who do not wear a hearing aid. It was recommended that the sub-pilot by UAB would try to focus on a relevant number of participants (> 25) with hearing aid to gather significant results for this group. It would be welcomed, if UAB tests could confirm the results for the group of hearing impaired users without hearing aid.

The content used is very important to obtain significant results. This means that for lab tests under controlled conditions it is strongly recommended to use "bad quality" content so an improvement in speech intelligibility is possible at all. Content needs to be carefully selected and evaluated with respect to the stereo mix.

The newly developed ad hoc methodology originally did foresee having an additional preference rating for each clip, but the results for the preference show a rather random distribution between the conditions/parts. The subjects often reported that it was too much for them to decide on a preference for one part (e.g. due to the length of the parts or the passed time since the beginning). Therefore, this preference rating did not provide any significant additional information. Hence, it is recommended to abandon the preference rating for future tests.

It is recommended to use 5.1 content for the sub-pilot. The (large) improvement in terms of speech intelligibility is more or less guaranteed. For stereo content, the CA generator can achieve small improvements, but not necessarily for every content.

7.2. Audio Description Test(s)

7.2.1. Sound mix

In October 2015 at the Sitges Film Festival UAB and TV3 performed some tests with secondary screens (Deliverable D4.1, chapter 7.1.2 **[1]**). One of the outcomes was users' complaints regarding sound mix for Audio Description. 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 (AD) sound levels. To this end the study has been designed.

Aim: find out if there could be some alternative way of AD sound edition process that improve user experience in the case of TV broadcasted AD to be mixed on the receiver.

Participants: 32 blind participants recruited through the Spanish user associations for the blind: ACCAPS, FESOCA, ONCE and ACIC.

Methodology: three clips lasting between 86 and 95 seconds were selected from Quentin Tarantino's 2009 film *Inglourious Basterds*. In all of the clips the sequences shown had a wide dynamic range in order to test the AD track in different sound conditions. The AD script was recorded only once for each clip. The recordings were edited in the same way and normalized to -23 LUFS **[3]**.

Once the recordings were ready, three different mixing processes were used to simulate three different scenarios:

1) *No adjustments* – AD track normalized at -23 LUFS; no volume adjustments made to either the AD track or the original soundtrack (OST);

2) *OST adjusted to AD* – AD track inserted at -23 LUFS; OST lowered by 3-4 decibels for clips 1 and 3, and by 2-3 decibels for clip 2, when AD was added;

3) *AD adjusted to OST* – OST left intact; the volume of the AD track manually adjusted to the volume fluctuations of the OST, i.e. accordingly lowered by 2-3 decibels or accordingly increased by 3-4 decibels.

The tests were carried out with a Samsung Galaxy Tab 3 tablet and Phillips RQH4 speakers connected by Bluetooth.

A questionnaire was drawn featuring nine statements to evaluate different sound perception criteria on a scale from 1 to 3, where 1=completely disagree and 3=completely agree.

The test was performed individually for each participant. They listened to each of the mixes in a randomized order and were asked to fill in the questionnaire after each clip.

The Spanish Blind Organization, ONCE, provided the facilities where the tests took place.

Conclusions: audio edition and mixing has an effect over comprehension and audibility of AD:

1) No adjustments of either the AD track or OST track can lead to comprehension and audibility problems (scenario 1);

2) The best way of guaranteeing a clear audibility of the AD track is to lower the OST by an average of 3 to 4 decibels, whenever AD is inserted (scenario 2);

3) Dynamic edition of the AD track shows less comprehension and more audibility problems compared to the mix with no adjustments (scenario 3).

7.2.2. Cultural Allusions & Intertextuality

User tests have been conducted in three countries: the United Kingdom, Italy and Spain.

Aim: measure how native and foreign audiences with and without vision impairments are engaged in the comprehension of the cultural and intertextual elements of audio described films.

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

Participants:

- *United Kingdom*: 15 sighted students from the University of Loughborough and seven visually-impaired users from the Royal National Institute for the Blind (RNIB) in London;

- *Italy*: 9 sighted students from the Trieste University and 10 visually-impaired users from the Unione dei Ciechi (Blind Union) in Udine;

- *Spain*: 11 sighted students from the Nebrija University and 11 visually-impaired users from the Spanish Association for the Blind "DOCE" in Madrid.

Methodology: *The King's Speech* (2010, dir. by T. Hooper) was selected for the tests. 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 also in its dubbed versions in Italian and Spanish. Similarly, audio described versions of the film existed in all three languages.

A questionnaire was drawn up for the test. It contained questions related to general comprehension (specific cultural and intertextual elements) and engagement. Questionnaires have been adapted for Italian and Spanish users (they have been translated from English into Italian and Spanish).

Users were shown the film with Audio Description and were asked to fill in the questionnaire after the screening. Visually-impaired users had the questions read out by sighted volunteers and dictated their responses.

Conclusions: the hypothesis hold true: native audiences (English) reported higher levels of comprehension and engagement than non-native audiences (Italian and Spanish).

- *Comprehension*: no great differences in the understanding of cultural elements emerged between native (British) and non-native audiences (Italian and Spanish), but, as expected, cultural elements were picked up more easily by the British users.

  While there was no huge divide in scores among sighted users, the differences were much more evident for the visually-impaired ones. Although comprehension seemed to have no effect on engagement, there was a noticeable lack of culturally-based knowledge on the part of the non-native audiences (Italian and Spanish).

- *Engagement:* British users greatly enjoyed the film and its particularly British "flavour"; Italian and Spanish users enjoyed the film, but were more reticent about its British "flavour", though none of them criticised this aspect directly.

7.3.Language Learning

7.3.1. Impact of Clean Audio on Language Acquisition

It is usual that second language students find difficult to understand the contents of a TV-recording or show. This is not only because they do not know the words, but also because they do not recognize those words due to the poor quality of the sound. That explains why understanding a phone call or a radio broadcast is normally harder than other instances when people first start communicating in a foreign language. Hence, the hypothesis is that Clean Audio will facilitate the first stages of this learning process by isolating the dialogues from the dispensable noise that hinders their proper absorption.

UAB has tested the impact of Clean Audio in language acquisition 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 Clean Audio generator).

Method: to validate the hypothesis that "cleaned" dialogues make it easier for language learners to isolate new items of vocabulary, UAB designed a recall and vocabulary questionnaire. If the hypothesis of Clean Audio helping with second language acquisition would be 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, Clean Audio could make a difference for people watching the original versions of films and TV shows (original meaning: not their mother tongue). Only material containing stereo audio mix was available for the tests.

Participants: 30 Spanish-speaking 12-year-old children who were tested with clips in Catalan.

Hypothesis: Clean Audio may improve language learner's experience by reducing the effort they must make in order to understand media in another language.

7.3.2. Clean Audio on Language Acquisition Test

UAB visited the Ben Arabí public high school in Cartagena (Spain) and carried out a vocabulary and recall test with 33 17 to 19-year-old students with no previous knowledge of the Catalan language. The test consisted of (see section **0**) three videos: one 30 second video of the TV show *Pop Rapid*, one minute and a half of the TV show *Vendelpla*, and one minute and a half of the documentary *Catalunya Experience*, all of them provided by TVC. The one minute and a half clips were divided into three parts, each processed in a different way: one unprocessed, two processed by the Clean Audio tool. The selection of the types of Clean Audio processing used in this test (processings A and C) were based on the results of the preference questionnaire for normal hearing users (see section **4.4.1**).

The class was divided in three groups of 12 students, and each group was shown the clips in the same order, but with them being processed in a different order, per the following scheme: group 1 will watch the clips in "C-A-hidden reference" order, group 2 in "A-hidden reference-C" order and group 3 in "hidden reference-C-A" order (please refer to section **11.2** for technical background).

There were one vocabulary and one recall question for the first show, and three vocabulary and three recall questions for the other two videos, each corresponding to a differently processed part, with a total of 18 items to be evaluated. The results on the questions for the first show, which was left unprocessed in all cases, are not included in the analysis, since it was shown with the only purpose of getting the students familiar with the sound of Catalan and the type of questions of the test. However, the students were not aware that these questions were not included in the process.

7.3.3. Results of the Language Acquisition Test

The results of the vocabulary and recall tests indicate that Clean Audio does not have an impact on second language acquisition, but also that the results of the test for the processed versions were worse when compared to the original unprocessed version (the hidden reference), see **Table 16**.

Table 16. Correct answers on the language acquisition test

Processing	Group 1	Group 2	Group 3	Total
A	24	9	8	41
C	5	10	13	28
Hidden reference	14	25	7	46

However, it is important to consider that these students had never been in contact with the Catalan language before. For this reason, we found it interesting to evaluate not only the vocabulary items that were totally correct, including their spelling, but also the ones that were phonetically close to the Catalan pronunciation. For example, we marked as correct the phonetically transcribed *trubat* for the right spelling *trobat*. Since Clean Audio aims to improve speech intelligibility, we consider spelling not to be significant for the determination of the viability of Clean Audio as a second language tool.

Table 17. Phonetically correct answers on the language acquisition test

Processing	Group 1	Group 2	Group 3	Total
A	27	13	8	48
C	15	11	16	42
Hidden reference	14	26	9	49

When adding the extra almost correct answers, we found that both the hidden reference and the Clean Audio processing A obtained almost the same amount of correct answers, see **Table 17**.

8. Ethical Issues and Data Protection

8.1. Ethical Requirements

The HBB4ALL project carried out tests with humans, as end users, in WP3, 4, 5 and 6. Access services were tested and trialled. Testing with end users was considered one of the project's strengths: the participation of those for who the services are mainly designed. This fulfils the UN CRPD "nothing about us without us"[6].

All tests were designed and complied with the relevant national, EU and international ethics-related rules and professional codes of conduct.

Universitat Autònoma de Barcelona (UAB), HBB4ALL coordinator, has an Ethical Commission on Human and Animal Research to supervise the experimentation on human and animal beings in compliance with the European directives 86/609/CEE, 91/628/CEE and 92/65/CEE. Given that other partners didn't have an Ethical Commission, and given the fact that the UAB commission fulfils all EU directives, it was decided that UAB would seek certificates for all tests.

There were three aspects which were requested permission by all tests:

(i) test design

(ii) informed consent and

(iii) privacy and data protection

In all tests the following issues were respected:

- Tests were planned, implemented and evaluated in a free and independent way.
- Contact with end users was conducted in a respectful way on an equal footing with all users. Especially people who are less competent must have increased attention by the test leaders.
- The tester must be informed honestly and give their consent. The communication must be adapted to the needs of users.
- The tests were anonymous and privacy was ensured.
- A pleasant atmosphere for the user needs was created, so that the test results were as free and objective as possible. To put the users under pressure regardless of the type (time, understanding, empathy) would distort the test results.

Forms used during the project:

1. Form to request permission (see Annex **0**)
2. Consent form (see Annex **11.5.2**)

[6] http://www.un.org/disabilities/documents/convention/convoptprot-e.pdf

3. Information to participants

8.2.Data protection

All data was anonymized. Also in HBB4ALL we took on board EU data protection policies following the European Directive 95/46 with date 24/10/1995, and also local policies such as the German Federal Data Protection Act (BDSG) or the Spanish Ley Orgánica de protección de datos 15/1999, and the different countries where tests were performed. Data was stored in an internal UAB server.

8.3.Sub-pilot specific issues and measures

With the aim of gathering quantitative usage data, TVC collects information about the navigation in its HbbTV service and the playback of the medias offered. This, and no other kind of data was collected during the Catalan Sub-Pilots: only the information was recorded that was strictly necessary to know the navigation and use of media items available to users in TVC's Catch-Up TV service. This data was collected anonymously and only referred to identifiers of session and user, this second identifier depending of the user TV device configuration about cookies behaviour. Moreover, the platform on which data were collected and analysed is a third party tool, Adobe Omniture. This tool always outputs its reports in an anonymized fashion - without the possibility of regaining detailed user information through the reports.

9. Conclusions

The final lab tests and field trail conducted in Germany indicate, that the Clean Audio processing developed by IRT can improve the speech intelligibility for TV programmes. The lab tests showed an improvement in speech intelligibility amongst the hearing impaired participants who do not wear a hearing aid. The field test coordinated by RBB confirmed, that the CA generator can achieve small improvements, but also confirmed that the effect of the processing strongly depends on the content and the audio mix. Further testing is suggested to further improve the understanding of the applicability of Clean Audio for TV viewers.

For the tests during the final phase of the project, stereo material was used as input to the Clean Audio tool. The results for 5.1 content had already shown that a significant improvement in terms of speech intelligibility can be reached with the developed CA generator. It is expected, that also for an operational service based on 5.1 material such an improvement can be obtained. This needs further verification in the field.

The tests conducted by UAB in Catalonia confirm the results from the German sub-pilot. Tests with the hard of hearing without hearing aids indicate that the speech intelligibility for Clean Audio is better, when compared to the original audio mix. At the same time, normal hearing test participants indicate, that they do not consider Clean Audio as a disturbing factor, but on the contrary may benefit also with respect to the speech intelligibility. Also the Catalonia tests were based on stereo material as input to the Clean Audio tool.

Based on the Audio Description user tests in the final project phase (carried out by UAB), recommendations have been made on the sound editing and mixing, as both have an effect on comprehension and audibility of AD. Additional tests confirmed the effect of cultural allusion and intertextuality in Audio Description, both increasing comprehension and engagement of users. These recommendations may be used by AD service providers and have been included as “best practices” in HBB4ALL deliverable D.2.6.2 **[4]**.

Both the Audio Description and Original Sound Track deployments by TVC have successfully been added as new features in its public HbbTV Catch-Up TV service “TV3alacarta”. Both audio features during the sub-pilots showed a stable acceptance of the new services. Based on the results, TVC has decided to maintain the AD and OST services as part of its HbbTV VoD service beyond the HBB4ALL project. Also, TVC is aiming at expanding this service and to bring these accessibility features to all platforms where TVC has presence. For a successful exploitation of the OST feature, a solution is required to technically limit the HbbTV application signalisation to the area of Catalonia. Large problems were encountered regarding the content rights, with respect to the distribution of OST versions via HbbTV VoD, currently resulting in a small amount of TV programmes that can be offered by TVC with this feature.

10. References

[1] D4.1 – Pilot-B Progress Report , HBB4ALL deliverable, 2014 (http://www.hbb4all.eu/wp-content/uploads/2015/03/D4.1-Pilot-B-Progress-Report.pdf)

[2] D4.2 - Pilot-B Solution Integration and Trials, HBB4ALL deliverable, 2015 (http://www.hbb4all.eu/wp-content/uploads/2015/03/D4.2-Pilot-B-Solution-Integration-and-Trials-2015.pdf)

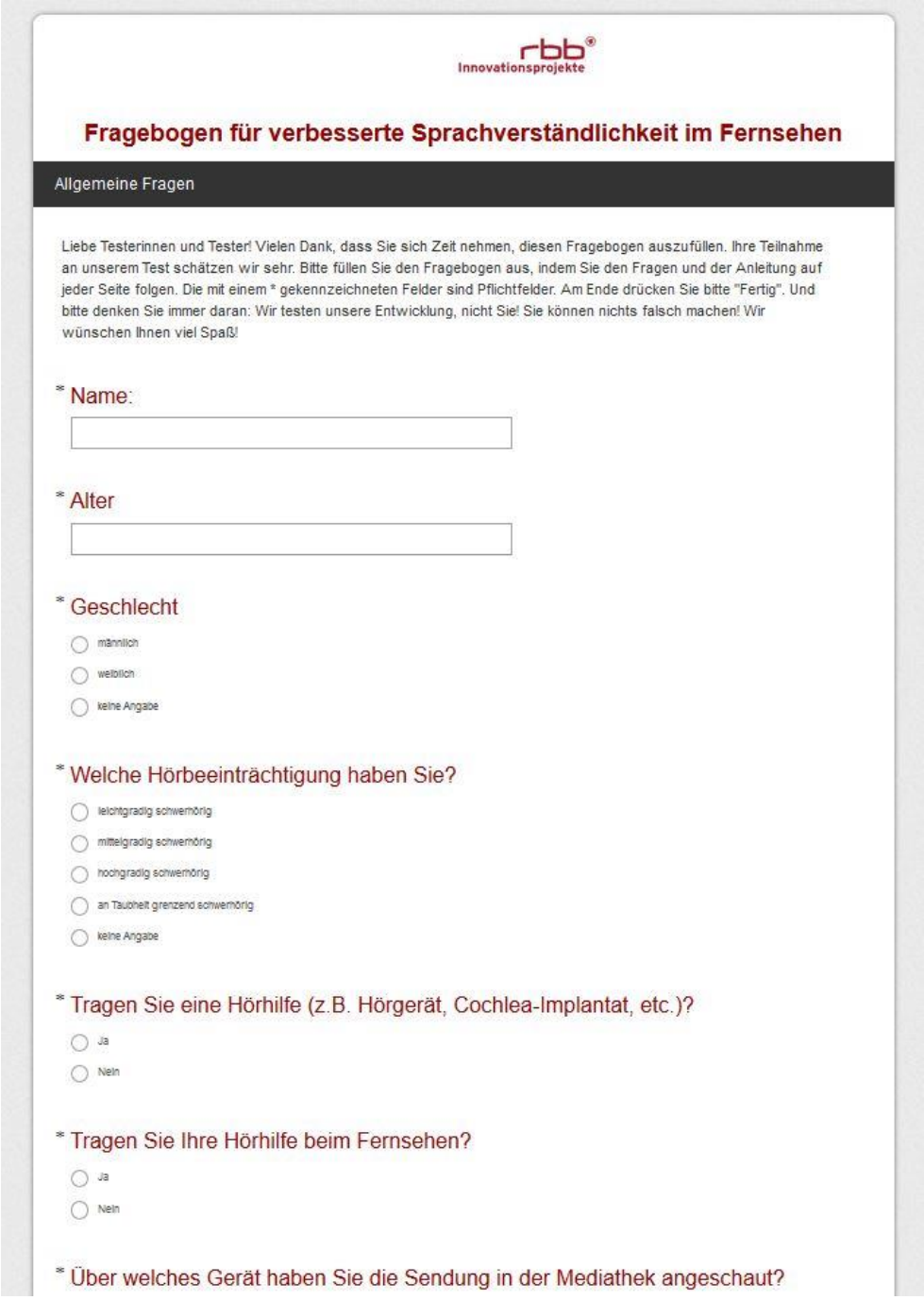
[3] Loudness Units Relative to Full Scale – More Info: https://en.wikipedia.org/wiki/LKFS

[4] D2.6.2 – Final Quality Metrics for TV access, HBB4ALL deliverable, to be published November 2016

11. Annexes

11.1. Clean Audio field test methodology

For the German pilot an online questionnaire was developed to collect a minimal amount of personal data that could have an effect on hearing and intelligibility of audio such as age, level of deafness, whether or not a hearing aid was worn and what type of equipment was used to play back the video and audio (see **Figure 24**). The second section of the questionnaire was designed to collect information about the difficulties the respondents had with the sound (specifically the speech intelligibility, see **Figure 25**). In this section they were asked to enter the problematic positions in the video based on the time code in the player and then describe the problems in their own words. The final part of the questionnaire included questions and scale to rate the overall sound quality of the video (**Figure 26**).



rbb Innovationsprojekte

Fragebogen für verbesserte Sprachverständlichkeit im Fernsehen

Allgemeine Fragen

Liebe Testerinnen und Tester! Vielen Dank, dass Sie sich Zeit nehmen, diesen Fragebogen auszufüllen. Ihre Teilnahme an unserem Test schätzen wir sehr. Bitte füllen Sie den Fragebogen aus, indem Sie den Fragen und der Anleitung auf jeder Seite folgen. Die mit einem * gekennzeichneten Felder sind Pflichtfelder. Am Ende drücken Sie bitte "Fertig". Und bitte denken Sie immer daran: Wir testen unsere Entwicklung, nicht Sie! Sie können nichts falsch machen! Wir wünschen Ihnen viel Spaß!

* Name:

* Alter

* Geschlecht

- männlich
- weiblich
- keine Angabe

* Welche Hörbeeinträchtigung haben Sie?

- leichtgradig schwerhörig
- mittelgradig schwerhörig
- hochgradig schwerhörig
- an Taubheit grenzend schwerhörig
- keine Angabe

* Tragen Sie eine Hörhilfe (z.B. Hörgerät, Cochlea-Implantat, etc.)?

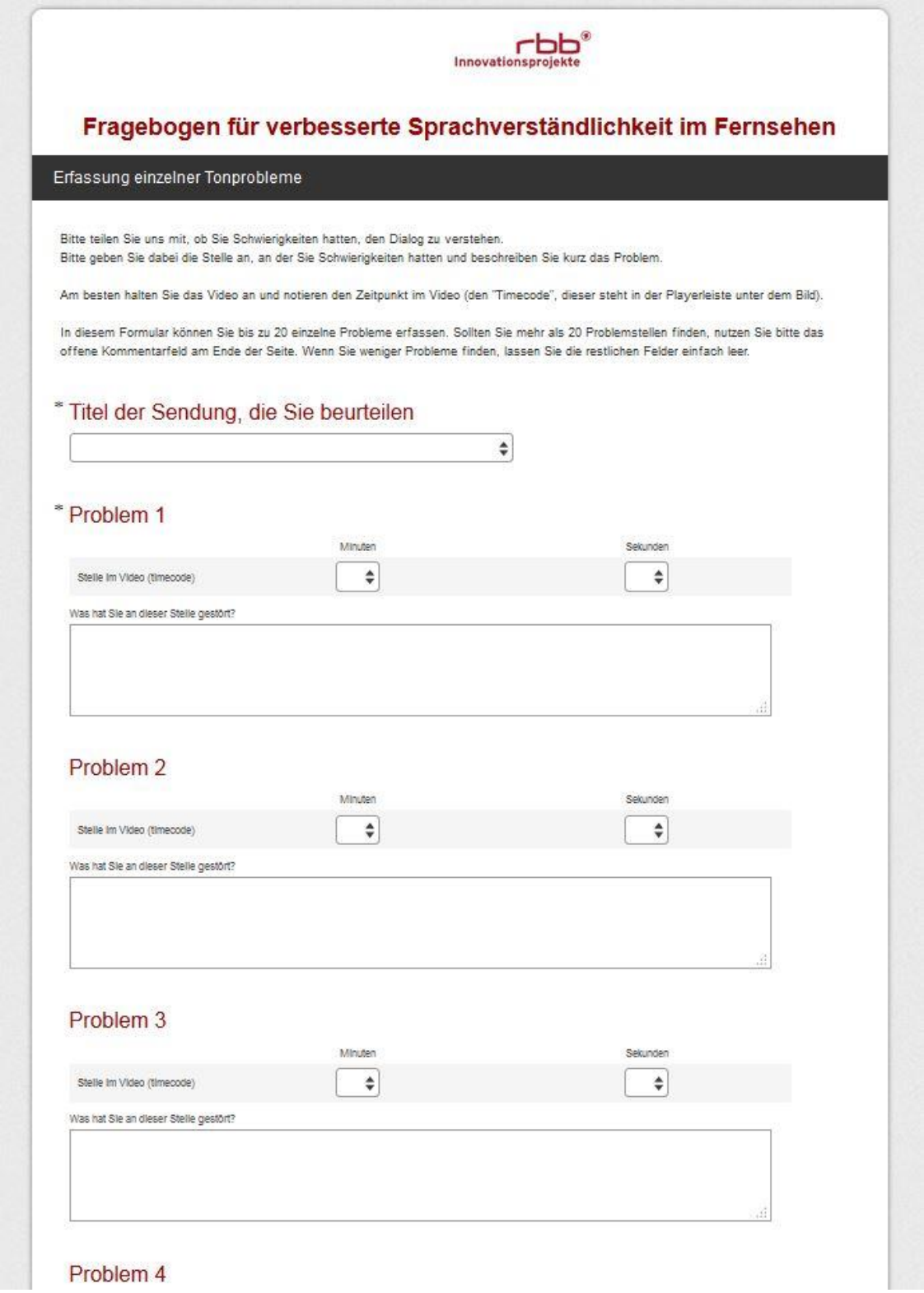
- Ja
- Nein

* Tragen Sie Ihre Hörhilfe beim Fernsehen?

- Ja
- Nein

* Über welches Gerät haben Sie die Sendung in der Mediathek angeschaut?

Figure 24. First page of online questionnaire for RBB Clean Audio Pilot



rbb Innovationsprojekte

Fragebogen für verbesserte Sprachverständlichkeit im Fernsehen

Erfassung einzelner Tonprobleme

Bitte teilen Sie uns mit, ob Sie Schwierigkeiten hatten, den Dialog zu verstehen.
Bitte geben Sie dabei die Stelle an, an der Sie Schwierigkeiten hatten und beschreiben Sie kurz das Problem.

Am besten halten Sie das Video an und notieren den Zeitpunkt im Video (den "Timecode", dieser steht in der Playerleiste unter dem Bild).

In diesem Formular können Sie bis zu 20 einzelne Probleme erfassen. Sollten Sie mehr als 20 Problemstellen finden, nutzen Sie bitte das offene Kommentarfeld am Ende der Seite. Wenn Sie weniger Probleme finden, lassen Sie die restlichen Felder einfach leer.

* Titel der Sendung, die Sie beurteilen

* Problem 1

Minuten Sekunden

Stelle im Video (timecode)

Was hat Sie an dieser Stelle gestört?

Problem 2

Minuten Sekunden

Stelle im Video (timecode)

Was hat Sie an dieser Stelle gestört?

Problem 3

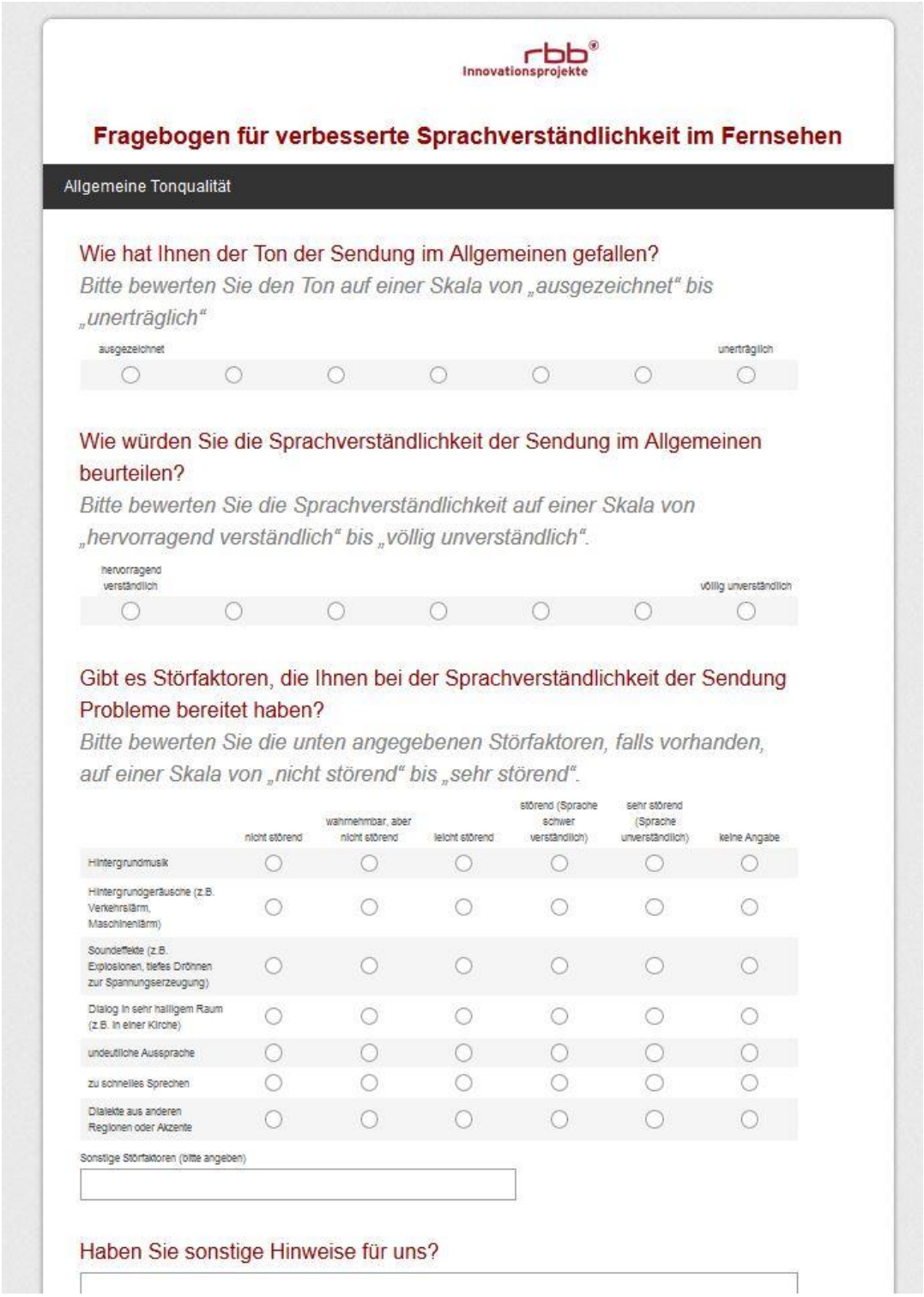
Minuten Sekunden

Stelle im Video (timecode)

Was hat Sie an dieser Stelle gestört?

Problem 4

Figure 25. Second page of online questionnaire for RBB Clean Audio pilot



rbb Innovationsprojekte

Fragebogen für verbesserte Sprachverständlichkeit im Fernsehen

Allgemeine Tonqualität

Wie hat Ihnen der Ton der Sendung im Allgemeinen gefallen?

Bitte bewerten Sie den Ton auf einer Skala von „ausgezeichnet" bis „unerträglich"

ausgezeichnet ○ ○ ○ ○ ○ ○ ○ unerträglich

Wie würden Sie die Sprachverständlichkeit der Sendung im Allgemeinen beurteilen?

Bitte bewerten Sie die Sprachverständlichkeit auf einer Skala von „hervorragend verständlich" bis „völlig unverständlich".

hervorragend verständlich ○ ○ ○ ○ ○ ○ ○ völlig unverständlich

Gibt es Störfaktoren, die Ihnen bei der Sprachverständlichkeit der Sendung Probleme bereitet haben?

Bitte bewerten Sie die unten angegebenen Störfaktoren, falls vorhanden, auf einer Skala von „nicht störend" bis „sehr störend".

	nicht störend	wahrnehmbar, aber nicht störend	leicht störend	störend (Sprache schwer verständlich)	sehr störend (Sprache unverständlich)	keine Angabe
Hintergrundmusik	○	○	○	○	○	○
Hintergrundgeräusche (z.B. Verkehrslärm, Maschinenlärm)	○	○	○	○	○	○
Soundeffekte (z.B. Explosionen, tiefes Dröhnen zur Spannungserzeugung)	○	○	○	○	○	○
Dialog in sehr halligem Raum (z.B. in einer Kirche)	○	○	○	○	○	○
undeutliche Aussprache	○	○	○	○	○	○
zu schnelles Sprechen	○	○	○	○	○	○
Dialekte aus anderen Regionen oder Akzente	○	○	○	○	○	○

Sonstige Störfaktoren (bitte angeben)

Haben Sie sonstige Hinweise für uns?

Figure 26. Final page of online questionnaire for RBB Clean Audio pilot

11.2. Clean Audio Lab Test Methodology

The new testing methodology for Clean Audio lab tests took into account the weak points of the previous method and should avoid:

- comparing a processed version of a test clip with the original and rating the difference between the two versions on a LIKERT scale (as this comparison seemed complicated for many users in the target group);
- repetition of a test clip, as the memory effect influences the ratings for a speech intelligibility test.

The new method that was designed by UAB and IRT therefore is based on listening to test clips only once and qualitatively rating the speech intelligibility for each clip (being it the original or a processed version).

In the 3rd and final Clean Audio lab test carried out by IRT, this new methodology was used for the first time (see section **7.1**). The content used for the test was chosen based on complaints from viewers about the bad intelligibility. From all potential clips, in a pre-evaluation a set of 25 clips was selected for the actual lab test, by a small group of hearing impaired people. The selected clips were at least 25 seconds long and had a similar audio composition over the whole duration (the same background signal(s) and speech).

Each selected clip was divided in 4 equally long parts (one for each test condition): one part was the original clip, the remaining three parts were processed with the Clean Audio generator, using different parameter settings to achieve three different Clean Audio versions. Following settings (being different for participants with and without a hearing aid) were used in the Clean Audio generator to generate the processed parts (explanation below):

Table 18. Clean Audio versions used in the 3rd Clean Audio lab test

CA-settings for group with hearing aid	**CA-settings for group without hearing aid**
Expander_noFilter_-18dB	Expander_noFilter_-18dB (Processing A)
noExpander_Filter-9dB_-18dB	noExpander_Filter-3dB_-18dB (Processing B)
noExpander_noFilter_HA-18dB	noExpander_Filter-9dB_-18dB (Processing C)

Explanation of the Clean Audio processing settings:

Each of the Clean Audio settings named in **Table 18** stands for one combination of the existing modules in the Clean Audio generator (please refer to D4.2 **[2]** for details to the implementation of the Clean Audio generator and its modules). "no" in the name implies that the respective module was not used for this combination:

- "Expander" means that the dynamic of the signal was changed by applying an expander to the output of the Center Cut algorithm (theoretically containing the speech).
- "Filter" stands for an additional frequency processing of the Center Cut output by amplification of the frequency band between 1600 Hz – 8000 Hz. This area is especially important for intelligibility of speech. The dB number (3 or 9 dB were used) after "Filter" stands for the gain value of the remaining (non-speech) frequencies (20 Hz – 1600 Hz and 8000 Hz – 20000 Hz).

- "-18dB" at the end of the name is the gain value for the CenterCut algorithm (used in all versions as we only used stereo content in the 3rd lab test).
- For the subjects with hearing aid, a special mode "HA" was offered in which the output of the CenterCut algorithm (theoretically the dialogue part) was panned to the left loudspeaker for playback, to avoid confusions of the hearing aid device with the Stereo phantom source.
- The settings for the group without hearing aid have been labelled (Processing A, B, C) to allow the cross reference of the test results (described in section **7.1.2)** with the Clean Audio settings presented in **Table 18**.

During the test, each clip was played to the participants where playback was paused after each of the four parts for a rating with respect to the intelligibility. The rating was a four-step qualitative scale with the following options (shown in German in **Figure 27**).

- I had **no problems** to understand the speech
- It was **a little exhausting** for me to understand the speech
- It was **exhausting** for me to understand the speech
- I **didn`t understand the speech at all**

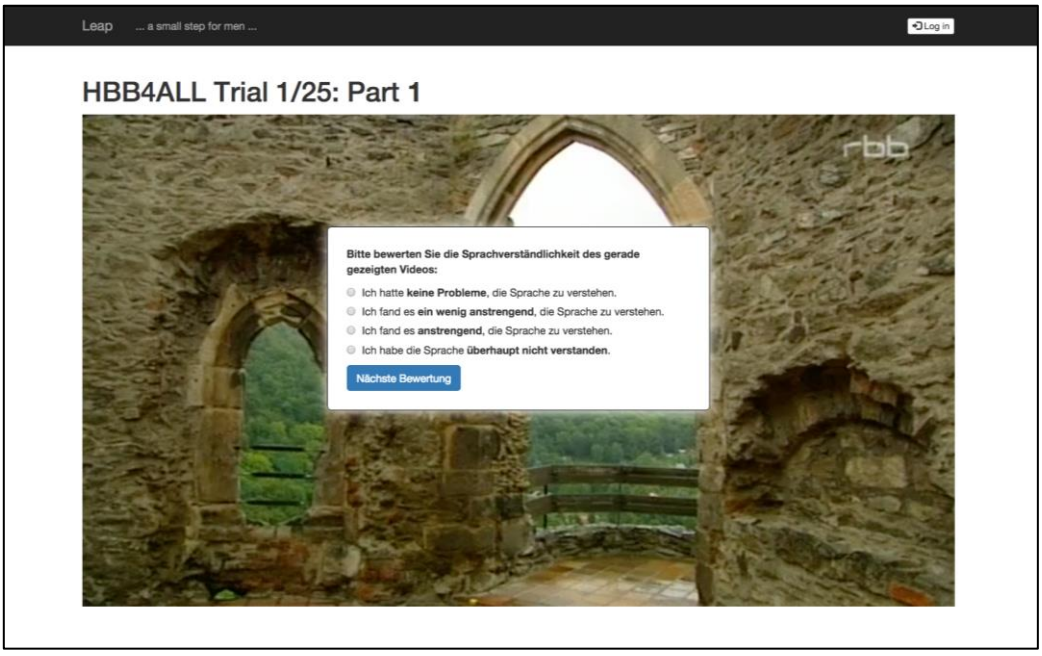


Figure 27. Clean Audio lab test methodology – rating of the clip parts on a four-step qualitative scale

At the end of the whole clip, after all ratings for the four parts had been given, the subjects also had to choose which of the four parts they preferred (showed in German in **Figure 28**).

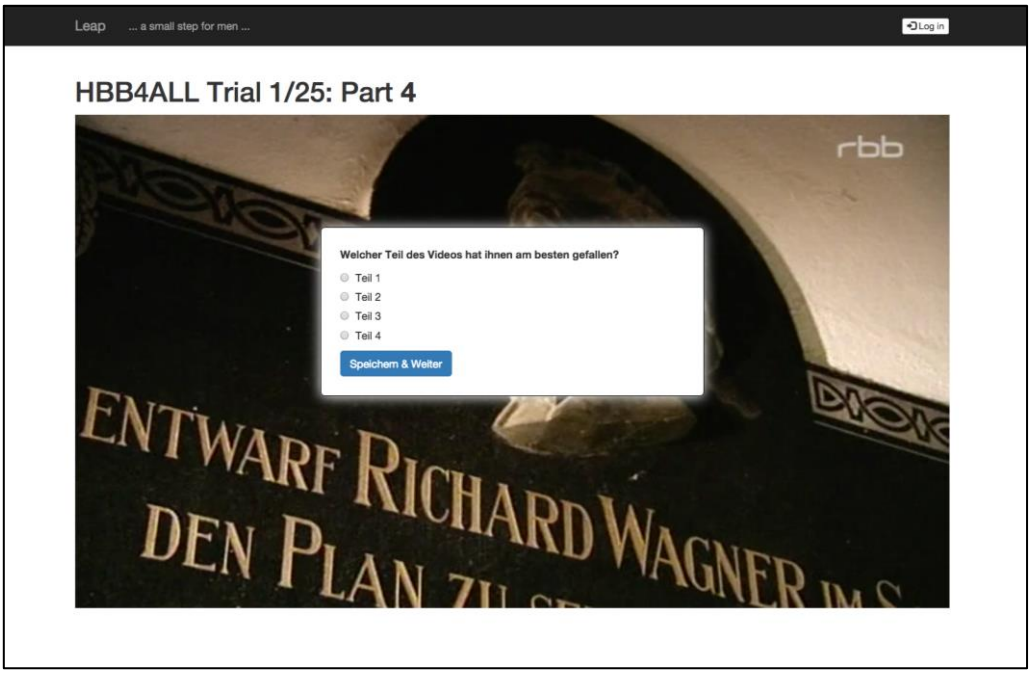


Figure 28. Clean Audio lab test methodology – selection of preferred part of the clip

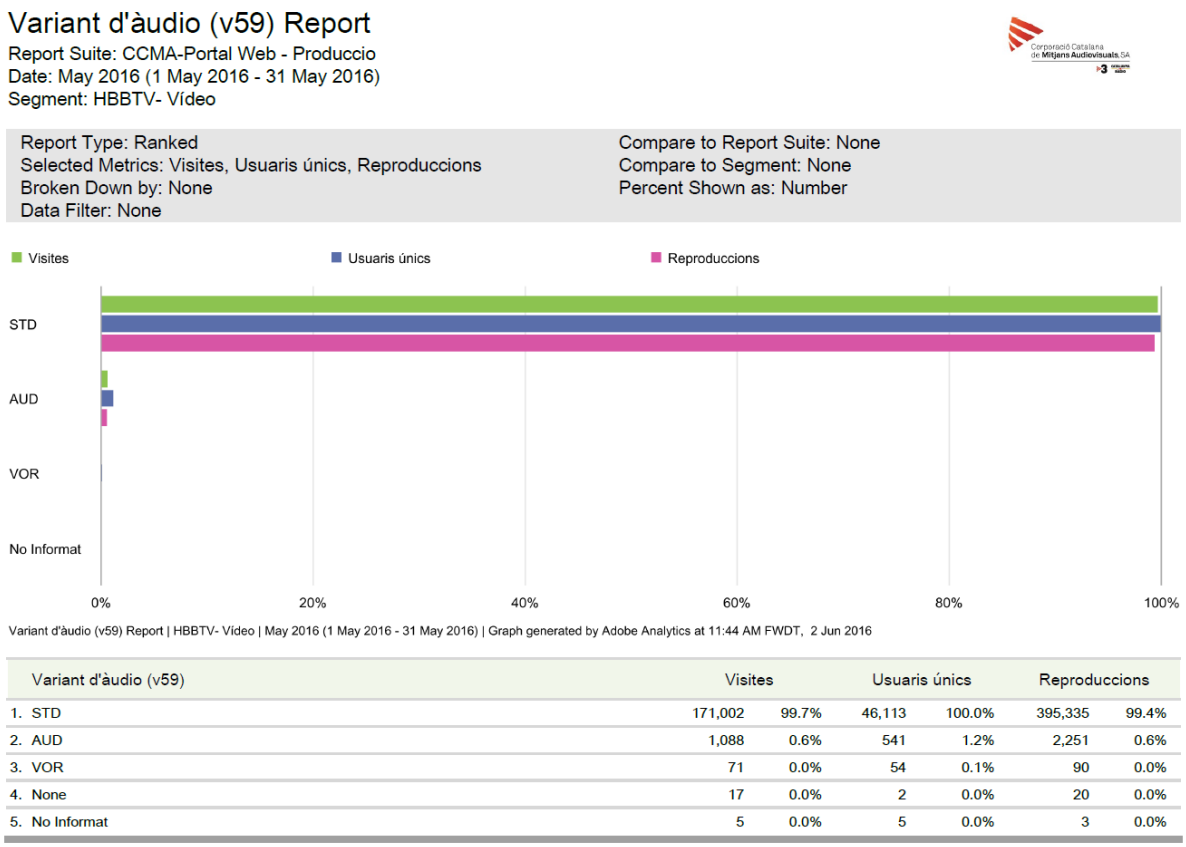
11.3. Overview of TVCs quantitative sub-pilot measurements

The TVC HbbTV application includes an internal system of labelling that allows collecting information on how the application is used. This process is done in the background without any visibility of it to the final users. The labels allow TVC to solve all sorts of questions concerning navigation and choices of the users. Every decision in navigation or play back are sent by the application to a platform specialized in collecting big amounts of data for posterior analysis. Concretely TVC choose Adobe Omniture Site Catalist environment for these collecting and analysis tasks.

For the HBB4ALL project a new label variable was created and it was collected whenever a user started a video playback. This specific label is called "audio variant" and can have the following values:

- "CLA": when the audio played is a Clean Audio version. Initially only one Clean Audio version has been foreseen. In case of more than one version of Clean Audio a new specific value for each should be created without quantity limitation.
- "NAR": when the audio played is an Explanation.
- "SIG": when the video is a Sign Language.
- "AUD": when the audio played is an Audio Description.
- "VOR": when the audio played is an Original Sound Track .
- "STD": when the audio played is a standard audio track, meaning a Catalonian (dubbed) audio track.

By evaluating the "audio variant" label, it can be determined what type of audio was played. This information, collected through the Adobe Omniture Site Catalist environment, can be accessed via Omniture web reports or, if preferred, can be downloaded as a csv data exchange format file.



Variant d'àudio (v59) Report
Report Suite: CCMA-Portal Web - Produccio
Date: May 2016 (1 May 2016 - 31 May 2016)
Segment: HBBTV- Vídeo

Report Type: Ranked
Selected Metrics: Visites, Usuaris únics, Reproduccions
Broken Down by: None
Data Filter: None

Compare to Report Suite: None
Compare to Segment: None
Percent Shown as: Number

Variant d'àudio (v59) Report | HBBTV- Vídeo | May 2016 (1 May 2016 - 31 May 2016) | Graph generated by Adobe Analytics at 11:44 AM FWDT, 2 Jun 2016

Variant d'àudio (v59)	Visites		Usuaris únics		Reproduccions	
1. STD	171,002	99.7%	46,113	100.0%	395,335	99.4%
2. AUD	1,088	0.6%	541	1.2%	2,251	0.6%
3. VOR	71	0.0%	54	0.1%	90	0.0%
4. None	17	0.0%	2	0.0%	20	0.0%
5. No Informat	5	0.0%	5	0.0%	3	0.0%

Adobe Marketing Cloud

Page 1

Figure 29. Adobe Omniture Site Catalist report example, aggregated "audio variant" values for month of May 2016

11.4. Questionnaire of the Clean Audio for Second Language Acquisition Test

The following language test was carried out by the 33 students in Ibn Arabí High School in Cartagena, Spain. The test includes vocabulary and comprehension questions about the clips that were shown to the students, such as fill in the blanks with new items of vocabulary that do not resemble Spanish, or direct questions about the content of the speech which could be answer with one to three words.

Vídeo 1	
Intenta rellenar la primera frase del diálogo: No ho entenc, de _____. Al taller hem fet moltes _____ aquests exercicis…	**¿Qué ha hecho Fede?**
Vídeo 2	
¿Cómo se dice "cabeza" en catalán?	**¿Qué representa la fachada de la Sagrada Familia?** - El centro: - El resto:
¿Cuándo se va a acabar la torre y cuánto va a medir? - Año: - Altura:	**¿Cómo se dice "cruz" en catalán?**
¿Qué forma dice el arquitecto que tienen las columnas? - Brazos - Árboles - Ninguna de las anteriores - No lo sé	**¿Qué quería mejorar Gaudí?**
Vídeo 3	
Intenta completar el diálogo: Has vist la Joana? _____ feliç. Sí, amb aquest _____ tan fabulós que ha _____...	**¿Cuánto tiempo lleva Pau en la discoteca?**
¿Qué problema ha tenido Martí?	**¿Cómo se dice "no te oigo" en catalán? ¿Y "no hay ningún problema"?**
¿Qué dice Martí que va a buscar y dónde? - Qué: - Dónde:	**¿Para qué apagan los móviles?**

11.5. Data Protection Agreements

11.5.1. Form to Request Permission

The following internal HBB4ALL form was sent to all partners requiring Ethical Commission approval, and it was then sent to UAB commission for approval. The form has the questions in Catalan but the replies are in English.

Títol Hybrid Broadcast Broadband for All

Breu descripció del projecte (3500 caracters)

The project HBB4ALL addresses media accessibility possibilities in the new hybrid broadcast-broadband TV (HbbTV) environment. To turn the accessibility vision into reality, HBB4ALL will address all relevant stakeholders and all components of the value chain. One of the prominent challenges of the coming years will be the multi-platform delivery of audio-visual content (anytime, anywhere, any device), be it a broadcast or an Internet service. Hybrid delivery platforms such as connected TVs and two-screen solutions enable a cost-efficient and convenient delivery of access services for those who need them. The elderly and people with various disabilities rely on subtitles, Audio Description, dialogue enhancement or sign interpretation. Customizing to personal preferences shall be possible within predetermined limits. The HBB4ALL project builds on HbbTV (from the existing versions 1.1.1 and 1.5 to the version 2.0 that is currently in development) as the major European standard for converged services and looks at both the production and service side. HbbTV provides a straight-forward specification on how to combine broadcast and broadband content plus interactive applications.
The project will test access services in various pilot implementations (from the definition to the operational phase) and gather implicit and explicit user feedback to assess the acceptance and the achievable quality of service in the various delivery scenarios.
HBB4ALL is elaborating pertinent guidelines, guides of good practice, metrics, and recommendations and will initiate campaigns to promote the project results. The results of HBB4ALL will be of worldwide relevance and will, through standardization bodies such as the ITU, also be publicized on a world-wide level. The overall objective of HBB4ALL is to become a major platform/player in the e Inclusion economy currently taking place, fostering the future market take-up while satisfying the diverse interests of all societal groups.

Area del procediment Antropologia

Objectius del procediment d'experimentació amb humans (3500 caracters)

Descriure els principals objectius que es pretenen assolir amb la realització d'aquest procediment d'experimentació

The objective of the user tests carried out in this project is to obtain quantitative and qualitative information about user's preferences and experience regarding accessibility services such as subtitling, Audio Description and Sign Language Translation.

The information gathered from users will be used to various ends:

- to determine the optimal parameters for the user's information processing
- to provide feedback to project partner's in order to customize and adapt existent accessibility technologies according to user's recommendations.
- to establish quality standards and guidelines for the presentation of different accessibility technologies to the public.

Archivo: No (Els fitxers adjunts han de ser com suport. Mai per substituir el text principal. Tots els documents han d'estar en format PDF)

Metodologia del procediment d'experimentació **(3500 caracters)**

Descriure breument la metodologia emprada justificant les dades, mostres biològiques i o respostes conductuals obtingudes de les persones sota experimentació

The general procedure of the user tests in this project is to present audiovisual content to the users such as films, clips or audio files, and verify user responses to a number of variables.

In order to do this, several qualitative and quantitative techniques are taken into account, namely administration of questionnaires, focus groups or interviews. Where needed, data will be recorded during the visualization phase (i.e. eye movements or time needed to perform a certain task).

Also, users will be required to perform certain actions like activating accessibility services (for example, activating subtitles or Audio Description) following previous instructions from the researcher.

Since this is a project aimed at addressing the needs of all the population, apart from people with no impairments, among the users there will also be the elderly and persons with hearing or vision loss. If a participant with a particular impairment needs assistance, a personal assistant will help him/her to perform the task in a given study. The participants with hearing or visual impairments will be recruited via official channels, i.e. by sending information to associations and institutions concerned with actions devoted for deaf and hard of hearing persons and blind and partially sighted persons and inviting them to our studies.

Before each actual test, a demographic questionnaire will be administered to gather background information on the participants.

Archivo: NO (Els fitxers adjunts han de ser com suport. Mai per substituir el text principal. Tots els documents han d'estar en format PDF)

Informació a les persones participants

S'annexa un full d'informació del projecte de recerca que inclou de forma entenedora els objectius de la investigació, els investigadors/res responsables i la forma d'obtenir fàcilment més informació?

Si.

Adjuntar Archivo: (veure al final)

S'annexa un full de consentiment informat signat per l'investigador/a i la persona en qüestió on queda clarament expressat que la participació és voluntària, que es podrà retirar en qualsevol moment sense donar explicacions, que disposa de la informació suficient i que en el cas d'estar sota tractament aquest no es veurà afectat de cap forma?

Adjuntar Archivo: (veure al final)

Compensació

Està previst algun tipus de compensació per la participació en el projecte?

No

Gestió i emmagatzematge de les dades obtingudes

Està prevista l'anonimització de les dades obtingudes?

Si

Està previst l'emmagatzematge de les dades en un servidor segur?

Si

Data collected in the course of the study will be stored on the Nebula server. Nebula is the space for collaborative work on documents as well as a repository for documents.

Feedback

Està prevista alguna forma de feedback a les persones participant un cop finalitzat el projecte?

Si

The participants will be given access to the project reports and academic articles produced in relation to the study.

Name of the project: HBB4ALL. Hybrid Broadcast Broadband for All

Aim of the tests

Methodology

Contact person

Forms were filled in in the many EU languages used for the tests: English, German, Polish, Italian, and Spanish.

11.5.2. Consent Form

As to informed consent, the partners were instructed to describe the procedure for obtaining the consent of persons, through a specific informed consent form. The informed consent form will be drafted in the language of the user, and will include standard features such as an explanation of the purposes of the tests, the expected duration of the test, a description of risks/discomforts/benefits to the subject etc., and will also indicate a contact person for pertinent questions. Since tests will take on board users of different sensorial disabilities, alternative communication channels (for example sign language or texts to be read by the visually impaired with Jaws) will be used.

Researchers will also describe the procedure the arrangements for protecting the confidentiality of personal data of the individuals concerned. If the researchers wish to retain the data for further research, they will have to ensure that the consent form mentions it and that the measures taken to encode or anonymize banked data are explained. In case only anonymized data will be retained, researchers will ensure adequate security for storage and handling of such data.

11.6. Technical implementation of Audio Description and Other Languages Sub-pilot

"TV3alaCarta" HbbTV application is a static HTML main page that consults (via an Ajax service) all the dynamic information referring to lists of medias available, layouts, banners, etc. For both the Audio Description and the Original Sound Track sub-pilots, the player asked for the specific metadata when the user wants to watch a specific program, also via the Ajax service. This metadata contains all necessary information (media files, subtitles, audio versions files, etc.) for the proper playback that is obtained from the Media Asset Management system.

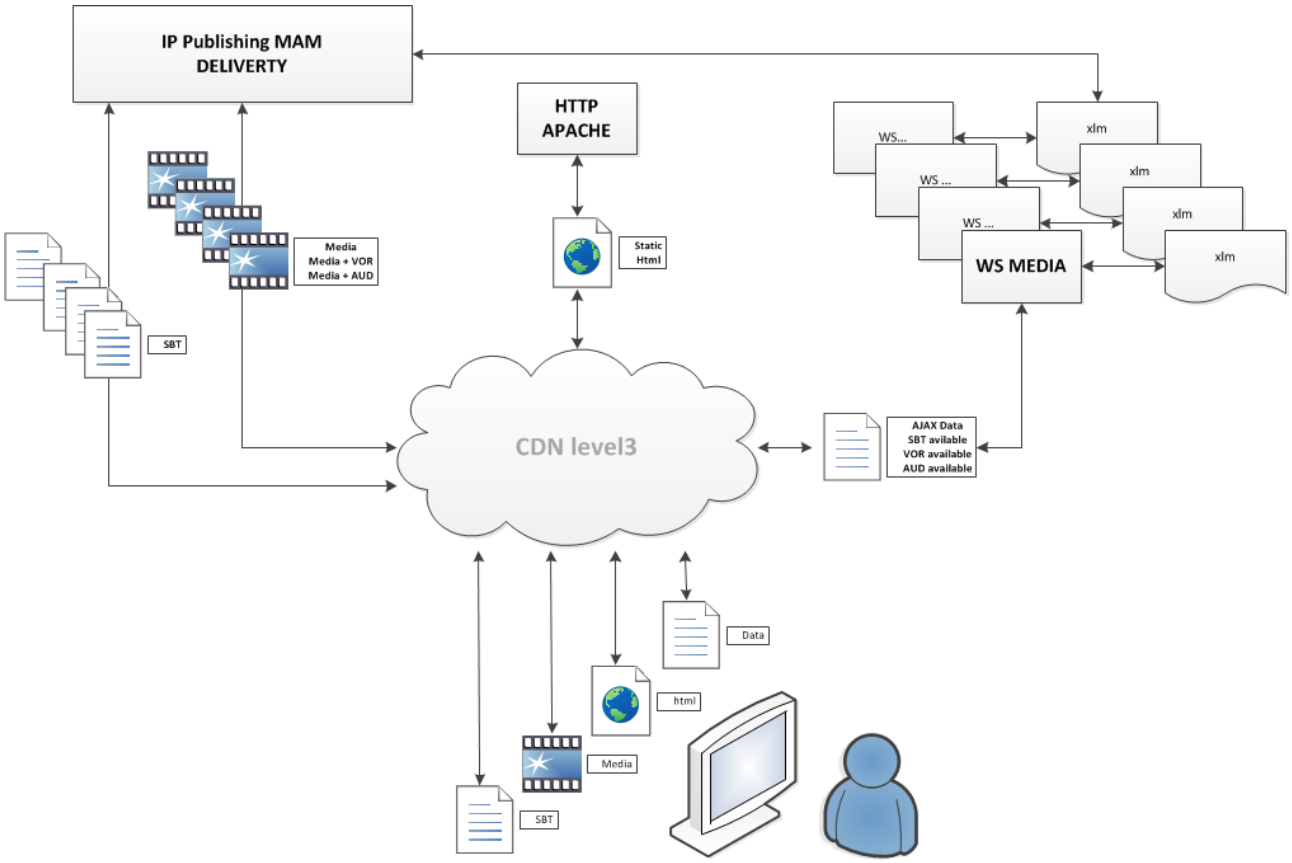


Figure 30. General TV HbbTV application scheme

The reply by the MAM to the player contains a URL to an MP4 file of the default asset. In the example shown in **Figure 31**, two variants are shown: AD and OST.

```
"media": {
        "geo": "TOTS",
        "format": "MP4",
        "url": "http://mp4-high-dwn.media.tv3.cat/8/3/1450495848038.mp4"
},
"variants": [{
        "id": "AUD",
        "nom": "Audio Descripció",
        "media": {
                "geo": "TOTS",
                "format": "MP4",
                "url": "http://mp4-medium-dwn.media.tv3.cat/1/9/1368856641291.mp4"
        }
}, {
        "id": "VO",
        "nom": "Versió original",
        "media": {
                "geo": "TOTS",
                "format": "MP4",
                "url": "http://mp4-medium-dwn.media.tv3.cat/1/9/1368856641291.mp4"
        }
}],
```


Figure 31. Reply of Media data information

In the case a variant is available (like AD or OST), the media player informs the users by adding a respective button on the media bar, see Figure 32.

Figure 32. OST and AD buttons in Media Bar

Initially, the media player starts playback of the default media asset, which is the asset with Catalan dubbed audio. If a viewer prefers to activate any variant, the media player switches to it and also saves the preference in a cookie. If the user does not change the state, the next time the media player finds a variant that matches the user's preference; the preferred variant will be played instead of the default media asset. To prevent conflicts there is a defined order of preference: Audio Description first, then Original Sound Track and original media asset.

Due to positive feedback in the sub-pilot, TVC is aware that must continue improving that functionalities, such as improve the "media" switching in a seamless way.