Final Requirements and Specifications for Pilot Demonstrators

Deliverable D2.4

CONTENT4ALL File ID:	C4A D2.4 Final Requirements and Specifications for Pilot Demonstrators.docx
Version:	1.0
Deliverable number:	D2.4
Authors:	Vaishnavi Upadrasta (HFC)
Contributors:	Wolfgang Paier (HHI), Marco Giovanelli (FIN), Richard Bowden (UNIS), Chathura Galkandage (UNIS), Robin Ribback (STXT), Sandra Böhm (HFC), Astrid Oehme (HFC)
Internal reviewers:	FIN
Work Package:	WP2
Task:	T2.1, T2.2, T2.3, T2.4
Nature:	R – Report
Dissemination:	PU – Public
Status:	Final
Delivery date:	14/10/2019

Version and controls:

Version	Date	Reason for change	Editor
0.1	30/07/2019	First draft	HFC
0.1	02/08/2019	First version	HFC
0.1	15/08/2019	Contributions from UNIS	HFC
0.1	16/08/2019	Contributions from HHI	HFC
0.1	19/08/2019	Contributions + comments from HFC	HFC
0.1	26/08/2019	Contributions + comments from FINCONS	HFC
0.2	27/08/2019	Second version	HFC
0.2	27/08/2019	Contributions from UNIS	HFC
0.2	27/08/2019	Contributions from STXT	HFC
0.2	29/08/2019	Second version update	HFC
0.2	29/08/2019	Additional contributions from UNIS	HFC
0.2	30/08/2019	Additional contributions from HHI	HFC
0.3	30/08/2019	Third version	HFC
0.3	04/09/2019	Review	FIN
0.3	13/09/2019	Comments addressed	HFC
0.3	17/09/2019	Second review + contributions from HHI	FIN
0.3	10/10/2019	Add contributions from UNIS	FIN
1.0	14/10/2019	Final Version	FIN

Acknowledgement: The research leading to these results has received funding from the European Union's Horizon 2020 Programme (H2020-ICT-2016-2, call ICT-19-2017) under grant agreement n° 762021.

Executive Summary

The goal of CONTENT4ALL project is to provide a low-cost solution for deaf people accessible TV-content, based upon the sign-interpreted version of content produced for the hearing. As the project is grouped into different phases, this deliverable about final use cases and requirements focuses on the CONTENT4ALL system. In Pilot 1 Demonstrator (Phase 1), a captured sign-interpreter is inserted into existing TV content as well as a database to collect, catalogue and analyse sign-language is established. Pilot 2 Demonstrator (Phase 3) proposes to develop an automatic sign-interpretation technology using a photorealistic sign-interpreter.

The purpose of this document is to report use cases and final requirements that describe the interaction of the components as well as the characteristics the system should demonstrate. All use cases and requirements are matched to the project objectives.

There will be some content throughout this document that traces back to D2.2 due to it remaining unaltered for D2.4. These sections will be comprised of a reference to the content of D2.2, with additional remarks wherever differences may have arisen, in order to keep this document simple and clear.

Table Of Content

List of Figures 5
List of Tables 6
1. Introduction 7
1.1. Purpose and objectives 7
1.2. Structure of the Document 7
2. CONTENT4ALL: Overview 8
2.1. System Architecture and Component Description 8
Pilot 1 Demonstrator (Phase 1) 8
Pilot 2 Demonstrator (Phase 3) 12
2.2. Reference Scenario 15
2.3. Pilot 1 Demonstrator (Phase 1) 15
2.4. Pilot 2 Demonstrator (Phase 3) 15
3. Use Cases 17
3.1. Transmission of subtitles from Broadcaster to Processing and Rendering Unit 17
3.2. Transmission of subtitles within the Processing and Rendering Unit 17
3.3. 3D Virtual Signer Language Translator in the Processing and Rendering Unit 18
3.4. Rendering of 3D Model from hand/body/face skeleton parameters 19
3.5. Content Preparation in the Processing and Rendering Unit for Media Distribution 19
3.6. Transmission of Signing 3D Model Rendered Video Stream and Main Broadcast Streams to the User Terminals 20
3.7. Business Models 21
3.8. Use Cases and Matching C4A Objectives 21
3.9. Risk Analysis 21
4. Requirements 23
4.1. Content Generation Requirements 23
4.1.1. Broadcaster Requirements 23
4.1.2. 3D Virtual Signer Language Translator Requirements 23
4.2. Rendering/Model Generation Requirements 24
4.2.1. Model Creation and Capture 24
4.2.2. Animation and Rendering 25
4.3. Networking Requirements 26
4.4. Encoding, Mixing and Streaming Requirements 27
4.5. User Requirements 28
4.5.1. In-depth requirements analysis using user-centred design approach 29

4.5.2. Prioritization of user expectations and final user requirements and specifications for Content4All system ... 33
4.6. Evaluation Tools ... 38
4.7. Business models ... 40
4.8. Requirements and Matching C4A objectives ... 40
5. Conclusion ... 41
6. References ... 42
Partner Short Names ... 43

List of Figures

Figure 1: CONTENT4ALL Overall Logical Architecture Phase1 ... 8
Figure 2: CONTENT4ALL Overall Architecture workflow ... 11
Figure 3: C4A Overall Logical Architecture Phase3 ... 12
Figure 4: C4A Overall Architecture Workflow ... 14
Figure 5: Process of T2.2 Requirement analysis ... 28
Figure 6: Categories and sub-categories of the semi-structured interview guide ... 30
Figure 7: Screenshots of the online questionnaire ... 33
Figure 8: Prioritization process - Delphi analysis ... 34
Figure 9: Summary of the established user requirements and out of scope requirements ... 37

List of Tables

Table 1: Final C4A Use Cases and Matching to Project Objectives 21
Table 2: Final C4A Risk Analysis 21
Table 3: Final C4A Broadcaster Requirements 23
Table 4: Final C4A 3D Virtual Signer Language Translator Requirements 23
Table 5: Final C4A Model Creation and Capture Requirements 24
Table 6: Final C4A Animation and Rendering Requirements 25
Table 7: Final C4A Networking Requirements 26
Table 8: Final C4A Encoding, Mixing and Streaming Requirements 27
Table 9: Final user requirements and specifications for Content4All system 36
Table 10: C4A Evaluation Tools Requirements 39
Table 11: Final C4A Requirements and Matching to Project Objectives 40

1. Introduction

1.1. Purpose and objectives

The goal of CONTENT4ALL project is to provide a low-cost solution for deaf people to access TV-content, normally produced for the hearing audience. Second goal is to create datasets and algorithms to enable automated sign-interpreted content creation. Different phases for achieving these goals are defined: In Pilot 1 Demonstrator (Phase 1), a captured sign-interpreter will be inserted into existing TV content as well as a database to collect, catalogue and analyse sign-language will be established. In Pilot 2 Demonstrator (Phase 3) an automatic sign-interpretation technology for deaf TV watchers will be developed, which is displayed via an animated photo-realistic human virtual signer, and exemplary demonstrated in a defined application scenario.

This document provides the use cases and final requirements, which are defined for CONTENT4ALL system for Pilot 1 (Phase 1) and Pilot 2 (Phase 3) Demonstrators. They address the interaction of the components as well as the characteristics the system should contain.

1.2. Structure of the Document

In terms of understanding the defined use cases and requirements of the CONTENT4ALL system Pilot 1 (Phase 1) and Pilot 2 (Phase 3) Demonstrators, this document initially presents an overview of CONTENT4ALL architecture and characteristics (chapter 2). In chapter 3 the use cases for CONTENT4ALL system are reported and assessed as to their matching CONTENT4ALL objectives. Furthermore, a risks assessment and mitigation plan for the use case implementation is provided.

In chapter 4, final requirements derived for CONTENT4ALL application are reported. As shown in the use cases in section 3, the requirements also cover broadly the main C4A system components as well as their interactions. These requirements are also matched to the project objectives. The deliverable ends with a conclusion (chapter 5) and references.

2. CONTENT4ALL: Overview

The goal of CONTENT4ALL (further abbreviated as *C4A*) is to develop a new approach for translating spoken TV content into sign-interpreted language, represented via a photo-realistic human virtual signer on a personalized stream for deaf people. Different phases are proposed to achieve this aim:

- Phase 1 will develop hardware and software to insert a captured human sign-language interpreter as a 3D model into existing TV content. Afterwards sign-language translation data will be collected, catalogued and analysed in order to create a database of subtitles, sign-language manual and non-manual for automatic sign-language translation (Phase 2). Learning-based sign classification and segmentation mechanism will be developed.
- Phase 3 proposes to develop an automatic sign-interpretation technology for the scenario of News and explore the potential of using the Phase 2 models for animating a photo-realistic human virtual signer.

This deliverable defines the use cases (chapter 3) and final requirements (chapter 4) of the C4A Pilot 1 (Phase 1) and Pilot 2 (Phase 3) Demonstrators. In terms of understanding, this chapter provides the system's architecture (components and their interaction), the reference scenario and the addressed Pilot 1 (Phase 1) and Pilot 2 (Phase 3) Demonstrators as an introduction.

2.1. System Architecture and Component Description

Pilot 1 Demonstrator (Phase 1)

The following high-level representation of the C4A Pilot 1 Demonstrator (Phase 1) system shows the logical composition and the logical interaction among the components: Broadcaster, Remote Studio, Processing & Rendering Unit and User Terminals. In Figure 1 we show the aforementioned components and their interaction/flows, which are necessary to achieve the project objectives:

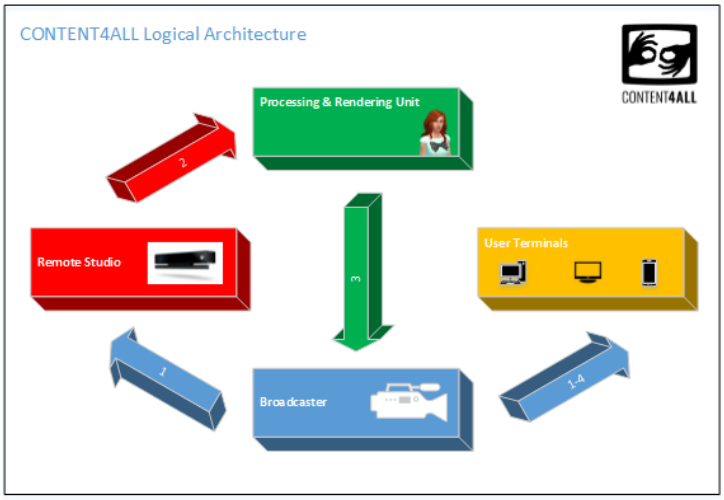

Figure 1: CONTENT4ALL Overall Logical Architecture Phase1

1. The main broadcaster video is streamed towards the Remote Studio and the User Terminals
2. The Remote Studio set-up receives the video stream and records the sign interpreter using Kinect and an additional camera. The stream of Kinect data along with the video of the Studio camera is streamed to the Processing & Rendering Unit component.
3. The Processing & Rendering Unit processes input data and reproduces the sign language through a 3D Photorealistic virtual signer.
4. Both video streams (original and mixed one) are sent to the users' terminals.

The Figure 2 describes the high-level architecture designed to satisfy the requirements of the Phase 1 of the CONTENT4ALL project. The reference architecture shows the different components and their interactions that will be implemented at the beginning of the project, distinguishing between components provided by the project (blue) and components already available in broadcaster premises (orange) that will be used for the project piloting. In this phase the automatic translation from video to signed content is replaced by a "remote studio" where a sign interpreter signs the video content and 3D rendering data are captured and transmitted to the "Photorealistic 3D Model Renderer" and transformed to a photorealistic 3D virtual signer.

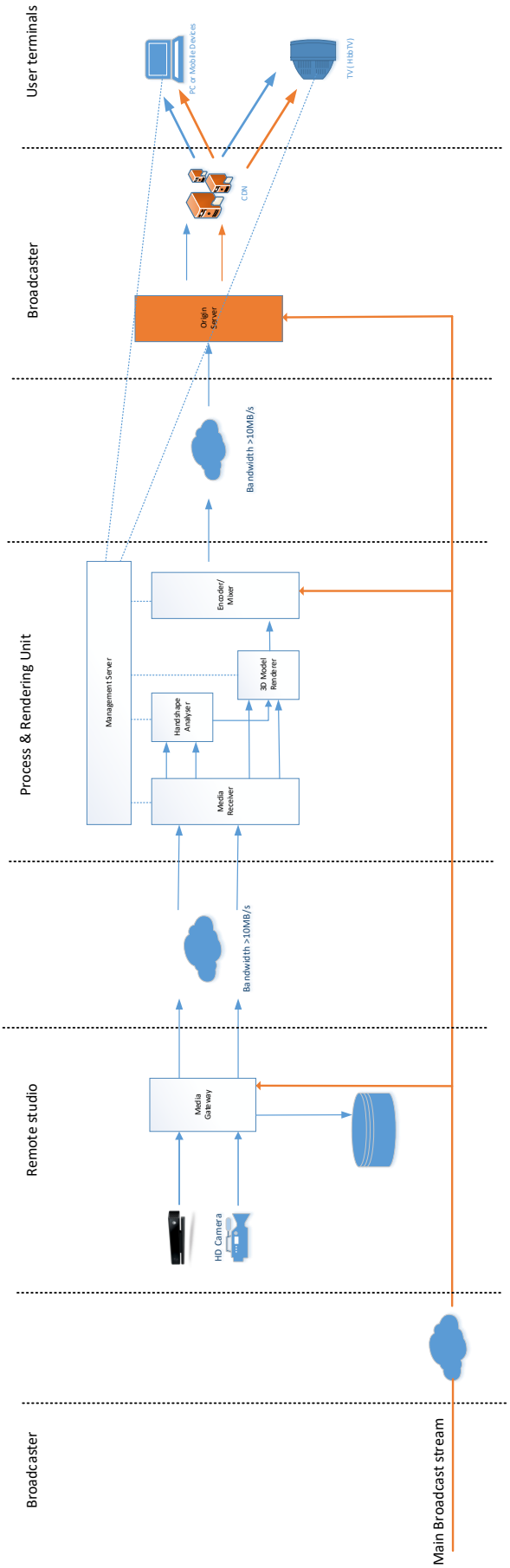
User terminals
PC or Mobile Devices
TV (Hbb TV)
CDN
Broadcaster
Origin Server
Bandwidth >10MB/s
Process & Rendering Unit
Encoder/ Mixer
3D Model Renderer
Handshape Analyser
Management Server
Media Receiver
Bandwidth >10MB/s
Remote studio
Media Gateway
HD Camera
Broadcaster
Main Broadcast stream

Figure 2: CONTENT4ALL Overall Architecture workflow

Logically the workflow can be divided in 4 components:

- Broadcaster
  - Main Broadcast stream
  - Origin Server
- Remote Studio
  - Kinect
  - HD Camera
  - Media Gateway
  - Storage
- Processing & Rendering Unit
  - Management Server
  - Media Receiver
  - Handshape Analyser
  - 3D Model Renderer
  - Encoder/Mixer
- User Terminals
  - PC or Mobile Devices
  - TV (HbbTV)

For further details, please refer to D2.3.

Pilot 2 Demonstrator (Phase 3)

A high-level representation of the C4A Pilot 2 Demonstrator (Phase 3) system architecture shows the logical composition and the interaction among the system's components: *Broadcaster*, *Processing & Rendering Unit* and *User Terminals*. In Figure 3 we show the aforementioned components and their interaction/flows, which are necessary to achieve the project objectives.

1. The main broadcaster video is streamed towards the user terminals.
2. The processing and rendering unit receive the main broadcaster video stream and reproduce the sign language through a 3D photorealistic model.
3. Both video streams (original and mixed one) are sent to the users' terminals.

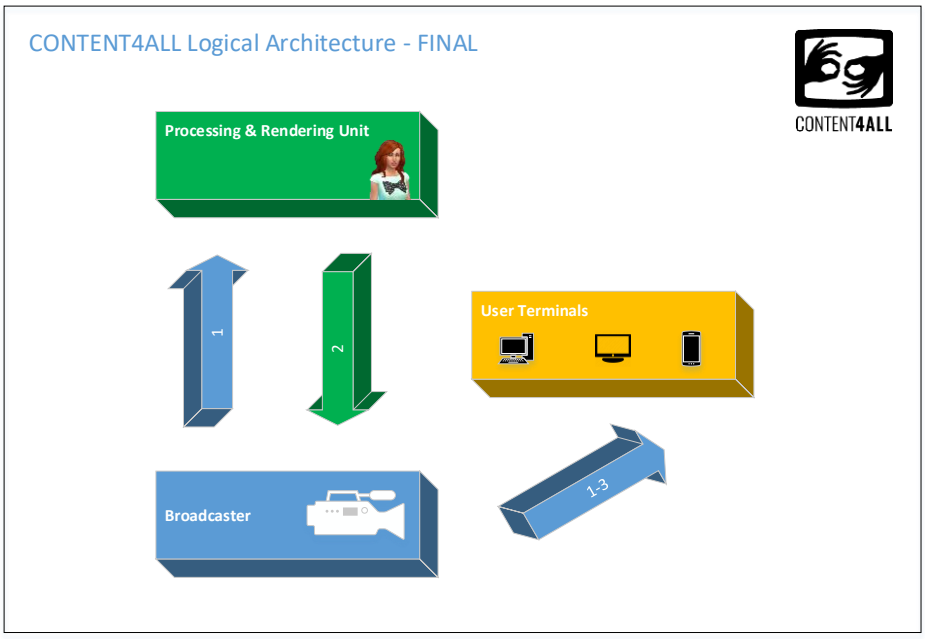

Figure 3: C4A Overall Logical Architecture Phase3

Figure 4 describes the C4A system's architecture more detailed. It shows the different components and their interactions that will be implemented at the beginning of the project, distinguishing between components provided by the project (blue) and components already available in broadcaster premises (orange) that will be used for the project piloting.

Main goals of demonstrator of at project's end, are:

- Automatic translation from video to signed content.
- 3D rendering data are transmitted to the photorealistic 3D model renderer and transformed to a photorealistic 3D human virtual signer.

The workflow can be divided in three components:

- Broadcaster
  - Main Broadcast stream
  - Origin Server
  - Media Gateway
- Processing & Rendering Unit
  - Management Server
  - Media Receiver
  - 3D Virtual Signer Language Translator
  - 3D Model Renderer
  - Encoder/Mixer

- User Terminals
  - PC or Mobile Devices
  - TV (HbbTV)

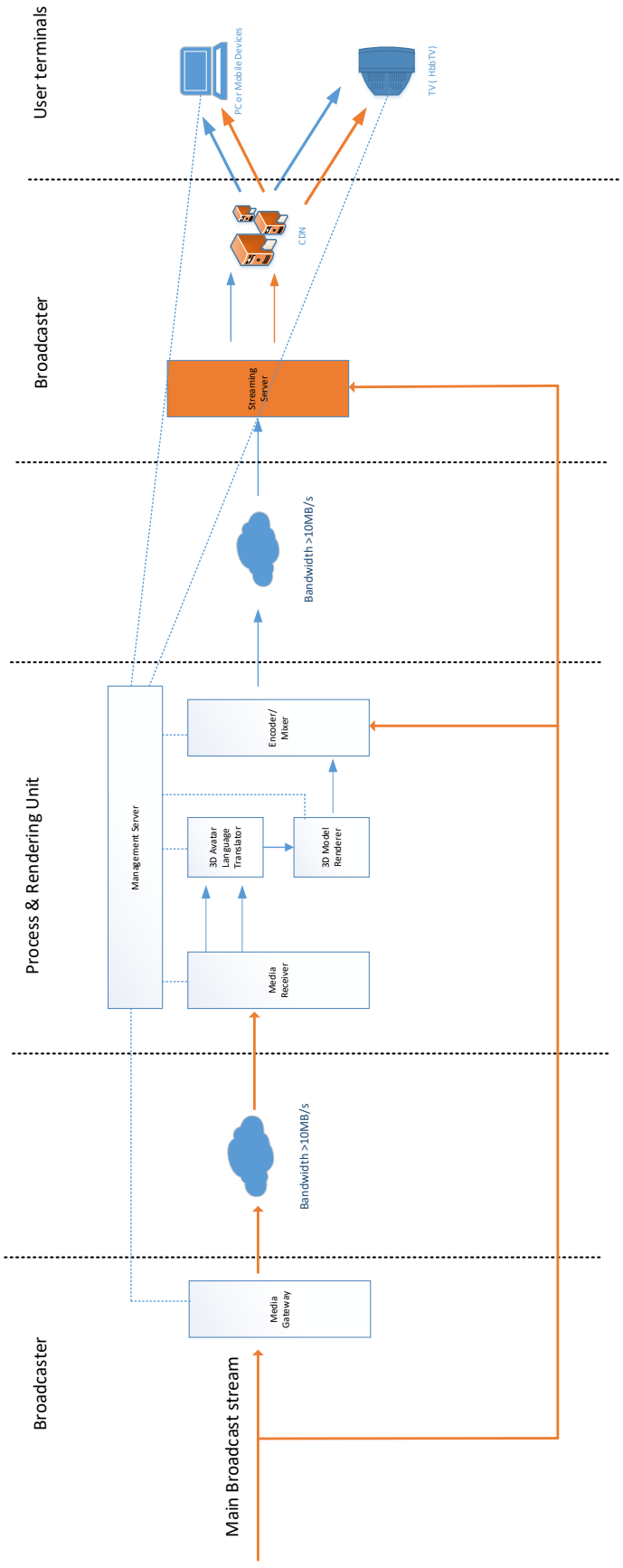

Figure 4: C4A Overall Architecture Workflow

For further details, please refer to D2.3.

2.2. Reference Scenario

The C4A framework will be tested during the transmission of *General News*, *Sport News* and *Weather News* content. These shows will be produced in two different languages: Swiss-German (STXT) and Flemish (VRT). Weather news will not be collected at VRTs premises, since now the broadcaster does not produce this kind of content for deaf people.

The aforementioned reference scenario will reflect the phases of the project. During the project two end-to-end demonstrations are planned, in order to demonstrate the technologies developed during the C4A project. In the first, all three TV shows will be produced and translated using the remote studio, while the second will focus only on weather news to demonstrate the results of the automated sign translation process.

Details of demonstrators will be included in D5.1 and D5.3, respectively, while a detailed description of the reference scenario, and differences between the phases can be found in D2.3.

2.3. Pilot 1 Demonstrator (Phase 1)

The first demonstration has been ready at M20 of the project. The Pilot 1 Demonstrator (Phase 1) followed this approach: C4A arranged two environments at the project's partner studios (STXT and VRT) in order to run and test the developed framework. When the sign interpreter at the remote studio starts signing in real time, a Kinect camera and depth sensor records his movements. In addition, facial expressions are recorded by the studio camera. Such information is sent to the processing and rendering unit, via the media gateway and forwarded to the handshape analyser and the photorealistic 3D model renderer components via the media receiver. The first component translates the input to a stream of metadata representing the intelligible format of handshapes for rendering. The metadata stream is the input to the photorealistic 3D model renderer along with the video and skeleton streams data provided by the sensors (aka Kinect and camera) placed in the remote studio. The model renderer component renders a 3D model. The 3D photorealistic renderer component adds the rendering of the handshape. The model's stream is forwarded to the mixer and encoder component that in turn assembles together the main video content (provided by the broadcaster) and the 3D model into a single stream that is further adapted (i.e. encoded in different formats).

The mixed stream is then sent to users through the origin server component. The translated content is streamed to different clients via the broadband connection. This content is streamed directly to televisions or set-top boxes that are HbbTV-enabled. Users can decide to switch from the original broadcast video to the broadband signed version directly using the remote control of their device. Since VRT does not to date support the standard HbbTV 2, a web application is implemented during the project, being able to receive and reproduce both the original and the translated contents. In doing so, the audience of possible users of the content is extended to anyone who owns a device with an internet browser, be it a PC or a mobile phone.

All data generated from remote studio are collected and stored into the storage systems already available in the broadcaster's infrastructure (Phase 2).

2.4. Pilot 2 Demonstrator (Phase 3)

For the Pilot 2 Demonstrator (Phase 3) the main change respect to Pilot 1 Demonstrator (Phase 1) is the removal of Remote Studio and the automatic generation of the 3D Virtual Signer from the Broadcaster Stream.

In fact, the Broadcaster Stream is sent to the processing and rendering unit, via the Media Gateway and forwarded to the 3D Virtual Signer Language Translator via the Media Receiver. The 3D Virtual Signer Language Translator component translates the text input (subtitles) to a stream of data that is sent to the photorealistic 3D model renderer. The model renderer component renders a 3D model. The model's stream is forwarded to the mixer and encoder component that adapts (i.e. encodes in different formats) the 3D model (eventually assembled together with the main video content provided by the broadcaster).

Similarly, to Pilot 1 Demonstrator (Phase 1), the stream is then sent to users through the origin server component. The translated content is streamed to different clients via the broadband connection. This content is streamed directly to

televisions or set-top boxes that are HbbTV-enabled, where (if not previously done) the 3D Virtual Signer is assembled together with the Broadcaster video. Users can decide to switch from the original broadcast video to the broadband signed version directly using the remote control of their device.

Since VRT does not to date support the standard HbbTV 2, a web application is implemented during the project, being able to receive and reproduce both the original and the translated contents. In doing so, the audience of possible users of the content is extended to anyone who owns a device with an internet browser, be it a PC or a mobile phone.

3. Use Cases

As already mentioned, the C4A system is built up of functional components which generate special content that is essential for the

- interactive content generation actions for the reference scenario, defined in D2.3, and recapped in section 2.2 above,
- the Pilot 1 (Phase 1) and Pilot 2 (Phase 3) Demonstrators, described in D2.3, and
- the project's overall objectives (see D2.3).

Use cases for the C4A system Pilot 1 Demonstrator (Phase 1) are already defined in D2.2. They remain almost unchanged, apart from the fact that the RGB and depth videos of the Kinect are directly processed in the Media Gateway in order to reduce bandwidth (for further details, please refer to the D2.3).

This section defines the main use cases which illustrate the interaction of the C4A system components for Pilot 2 Demonstrator (Phase 3) in terms of the project's objectives. All use cases will be described with pre-conditions, post-conditions and at least one main success scenario. Pre-conditions describe the states of the C4A system that have to exist before the use case can start. Post-conditions deal with the states of the system after the use case has been successfully completed. The main success scenario describes what the Pilot 2 Demonstrator (Phase 3) will (or will not) achieve, as a result of the respective use case.

The main functionalities of C4A system are covered and linked to the actors (mainly system components) that are addressed.

3.1. Transmission of subtitles from Broadcaster to Processing and Rendering Unit

The Media Gateway intercepts the Main Broadcaster Stream and then subtitles are transmitted to the Processing and Rendering Unit for 3D Virtual Signer Language Translation (see D2.3). The functionalities of Processing and Rendering Unit depends on Media Gateway providing a continuous stream of data. Therefore, use cases are defined for content preparation, encoding, muxing and streaming as given below.

Use case 1.1: Main Broadcaster Stream is intercepted by the Media Gateway that extracts and streams subtitles to the Media Receiver at the Processing and Rendering Unit.

Pre-conditions:

- Input port of the Media Gateway are open and are listening to the streams of the Broadcaster
- TCP server output port of the Media Gateway is successfully connected to the Media Receiver
- A continuous stream of raw transcription text data is available in the encoder buffer ready to be streamed.

Post-conditions:

- A continuous stream of subtitles is available in the Media Receiver.

Main success scenarios:

1. Media Receiver receives a continuous stream of subtitles.

3.2. Transmission of subtitles within the Processing and Rendering Unit

Once the subtitles are available in the Media Receiver, they need to be made available for the internal components in the Processing and Rendering Unit. Therefore, use cases to facilitate these transmissions are defined as follows.

Use case 2.1: Subtitles are transmitted from the Media Receiver to the 3D Language Translator.

Pre-conditions:

- Media Receiver in the Processing and Rendering Unit receives the subtitles.
- TCP server ports are available in the 3D Language Translator to receive the subtitles.

- Media Receiver is successfully connected to the 3D Language Translator.

Post-conditions:

- 3D Language Translator receives a continuous stream of subtitles.

Main success scenarios:

1. 3D Language Translator is running continuously to output hand-body-face skeleton parameters.

The 3D Language Translator component analyses the subtitles and determines the hand-body-face skeleton parameters, which are provided to the 3D model renderer in order to be rendered into the 3D model. The following use case is defined to transmit these parameters to the 3D model renderer.

Use case 2.2: The hand-body-face skeleton parameters are transmitted from the 3D Language Translator to the 3D model renderer.

Pre-conditions:

- 3D Language Translator determines the hand-body-face skeleton parameters.
- A JSON stream of hand-body-face skeleton parameters are available.
- 3D Language Translator is connected to the 3D model renderer through a TCP socket to transmit the data.

Post-conditions:

- Hand-body-face skeleton parameters are available at the 3D model renderer.

Main success scenario:

1. 3D Language Translator outputs a hand-body-face skeleton parameters to the 3D model renderer.
2. 3D model renderer receives and renders the hand-body-face skeleton parameters.
3. 3D model renderer continuously renders and outputs the rendered 3D model images at 25fps.

Once the 3D model renderer receives the hand-body-face skeleton parameters, it renders the 3D model. Once 3D model images are rendered, they need to be eventually mixed with the original broadcast stream (for backward compatibility) and encoded to be streamed to the users. Hence, the rendered 3D model images need to be made available in the mixer/encoder module in the Processing and Rendering Unit, for which a relevant use case is defined as below.

Use case 2.3: The rendered 3D model image is transmitted from the 3D model renderer to the mixer/encoder for encoding the stream for distribution.

Pre-conditions:

- Encoding/mixing module is running with input ports listening to the 3D rendered image.
- 3D model renderer is successfully connected to the encoding/mixing module.

Post-conditions:

- 3D model images are available in encoder/mixer at 25fps.

Main success scenario:

1. Encoder/mixer continuously encodes the 3D Virtual Signer video in multiple profiles and output them as MPEG-TS.

3.3. 3D Virtual Signer Language Translator in the Processing and Rendering Unit

The Pilot 2 Demonstrator (Phase 3) focuses on rendering a 3D model from subtitles into the broadcast stream in order to facilitate the sign language translation of the broadcast content. Therefore, the 3D Virtual Signer Language Translator

component is proposed to provide the 3D model renderer with the hand-body-face skeleton parameters. In this context, the following use cases are defined to describe the functionalities of the 3D Virtual Signer Language Translator.

Use case 3.1: The sign translator converts incoming subtitles to a stream of 3D Virtual Signer parameters which represents a sign translation of the sentence.

Pre-conditions:

- The Language Translator receives a subtitle from the Media Receiver

Post-conditions:

- A series of hand-body-face parameters are available for the 3D renderer.

Main success scenario:

1. The sequence of parameters represents an accurate translation of the sentence.

3.4. Rendering of 3D Model from hand/body/face skeleton parameters

For Pilot 2 Demonstrator (Phase 3), an important part of the rendering module is the animation of hand, body and face. The hand-body-face skeleton information, which is provided by the 3D Virtual Signer Language Translator, must be transferred onto the 3D model's skeleton such that the 3D model is performing the desired body movements. Therefore, a realistic animation must be created that allows natural movement which is defined in the use case below.

Use case 4.1: Hand/body/face animation from skeleton and face parameters.

Pre-conditions:

- The 3D model renderer receives the expression data from the 3D Virtual Signer Language Translator.

Post-conditions:

- The C4A 3D model's new pose reflects the pose, handshape and facial expression provided by the received expression data.

Main success scenario:

1. The C4A 3D model shows the hand-body-face pose/movements.

3.5. Content Preparation in the Processing and Rendering Unit for Media Distribution

For Pilot 2 Demonstrator (Phase 3) the mixing of original video and signing Virtual Signer will be done only for the devices that does not support it at client side. The Processing and Rendering unit will generate and encode two video streams: one mixed and one containing only the 3D Virtual Signer. The video streams are encoded with codecs (H.264/AVC) supported by selected client devices (HbbTV, pc, selected smartphones) and in different profiles, and send the output to the origin server in a proper format (MPEG-TS). Different use cases are defined.

Use case 5.1: The 3D Virtual Signer video stream is encoded into mpeg-ts container.

Pre-conditions:

- Input port of ffmpeg encoding instance is open and keeps on listening to the stream from the 3D model renderer.
- Output port of ffmpeg encoding instance is successfully connected to the origin server.
- Input and output buffers of the encoders are maintained at sufficient levels.
- Appropriate encoding setting, bit rate, and quality requirements are defined for the encoding instance.

Post-conditions:

- The H.264/AVC compressed video streams are available for streaming from the mixer and encoder to the origin server.

Main success scenario:

1. The origin server receives a continuous video stream of the 3D model
2. The origin server delivers the video stream to the user terminals.

3.6. Transmission of Signing 3D Model Rendered Video Stream and Main Broadcast Streams to the User Terminals

In order to transmit the 3D Virtual Signer video to client applications on the different devices, the origin server will package the video stream in an HTTP-based streaming protocol supported by the client devices. Once the video is packaged, the client applications can download the video stream (in small chunks) directly from the origin server (possibly via a Content Delivery Network that caches the chunks for increasing overall performances).

Use case 6.1: The HbbTV app downloads the 3D Virtual Signer video stream from the origin server.

Pre-conditions:

- The video feed from mixer and encoder is available.
- The origin server is packaging the stream in a suitable protocol.
- HbbTV app is ready to request video stream chunks.

Post-conditions:

- The mixed video stream chunks are received by the HbbTV app.

Main success scenario:

1. The mixed video stream is received and showed on user's terminals.

3.7. Business Models

Different business models for the C4A system Pilot 1 Demonstrator (Phase 1) are defined: for further information please see D2.2.

For what concerns the Pilot 2 Demonstrator (Phase 3), since it is a proof-of-concept, Business models will be analysed at a more advanced exploitation stage.

3.8. Use Cases and Matching C4A Objectives

For what concerns the Pilot 1 Demonstrator (Phase 1) matching C4A Objectives, please refer to D2.2.

The use cases of the Pilot 2 Demonstrator (Phase 3) that are identified in previous sections fulfil the following C4A main objectives, listed in the following table.

Table 1: Final C4A Use Cases and Matching to Project Objectives

Project Objectives	Matching use cases
1. 3D model generation and animation	4.1, 4.2
2. Signal processing for sign-language translation	3.1
3. Building a proof-of-concept demonstrator	1.1, 2.1, 2.2, 2.3, 5.1, 6.1

3.9. Risk Analysis

Risk Analysis for the C4A system Pilot 1 Demonstrator (Phase 1) is already defined: for further information please see D2.2.

For what concerns Pilot 2 Demonstrator (Phase 3), several risks and challenges regarding the realisation of use cases due to dependencies among components, tasks, interactions and inter-relationships do exist. The following table lists risks that may arise during system development associated mitigation strategies.

Table 2: Final C4A Risk Analysis

Risk	Likelihood	Impacting use cases	Mitigation plan
Translation accuracy is poor in the 3D Virtual Signer Language Translator	Medium	3.1	The domain of discourse can be further reduced in complexity.
3D model shows unnatural body poses	Medium	4.1	Improve the generated 3D skeleton with smoother animations and more natural limbs movement
The Model does not display the desired facial expression	Medium	4.1	Increase of training data for model generation by adding facial expressions that cannot be reproduced.

Risk	Likelihood	Impacting use cases	Mitigation plan
Packet losses due to buffer overflow or network overflow	Low	1.1, 2.3	Use TCP sockets for data exchange, which guarantees no packet losses.
TCP ports become unavailable for communication	Low	1.1, 2.1, 2.2, 2.3	Perform TCP port availability checks before the program initialisation.

4. Requirements

Requirements for the C4A system Pilot 1 Demonstrator (Phase 1) are already defined in D2.2. They remain almost unchanged, apart from the fact that the RGB and depth videos of the Kinect are directly processed in the Media Gateway in order to reduce bandwidth (for further details, please refer to the D2.3).

As concerns the Pilot 2 Demonstrator (Phase 3), in this section the requirements are specified which define the characteristics of the C4A system to be satisfied. The requirements face the wide scope of user requirements, which can, also like use cases, be assigned to the main C4A components.

4.1. Content Generation Requirements

It is important to produce and deliver quality content that will then be used during the creation phase of the automatic sign language translation model and the generation of the photorealistic 3D model. The higher the data quality is, the better the product model will be. We can rely on the standard components constituting the infrastructure of a broadcaster to guarantee a high level of quality and reliable broadcast of content.

This section derives requirements for the sign data collection for training and validating of the sign language recognition and translation work to be carried out in WP4 as well as requirements of the broadcasters to generate the data. The following requirements and methodologies to fulfil them are listed below in sec. 4.1.1 and 4.1.2.

4.1.1. Broadcaster Requirements

Table 3: Final C4A Broadcaster Requirements

Requirement ID	Req. description	Methodologies followed to fulfil the requirement
REQ1.1	The mixed video content stream must be sent to the user's terminals.	To stream the video content through the user's terminals we leverage on existing standard Broadcaster systems like origin server and possibly content delivery network system, when needed.

4.1.2. 3D Virtual Signer Language Translator Requirements

Table 4: Final C4A 3D Virtual Signer Language Translator Requirements

Requirement ID	Req. description	Methodologies followed to fulfil the requirement
REQ1.2	The subtitles must be sent to the 3D Virtual Signer Language Translator.	The Media Gateway extracts subtitles form the Main Broadcast Stream, send them to the Media Receiver that forwards them to the 3D Virtual Signer Language Translator.

4.2. Rendering/Model Generation Requirements

This section contains a list of requirements that need to be considered for the 3D model rendering and model creation. Parts of this list are directly deduced from user requirement research and transformed into technical requirements, as this is highly important for the success of the developed technology. Additionally, we listed requirements that need to be considered during the model creation process to ensure the high quality of the captured data. Finally, we considered requirements that are important in order to achieve the proposed technological targets of this project.

4.2.1. Model Creation and Capture

Based on the user requirements research the 3D model should be a human adult and have a realistic appearance. The 3D model should also be able to display realistic facial expressions and perfect emotions. Since a high level of realism is hard to achieve with geometry-based animation alone, it will be necessary to create a set of dynamic textures that are able to capture fine facial movements.

In order to perform the geometry and texture analysis we found some additional requirements that apply to the captured sign language translator. These requirements will ensure that the quality of the captured geometry, textures and movements is high and free of artefacts (e.g. tracking errors, wrong geometry and texture information for example caused by glasses).

The derived requirements are listed in the following table.

Table 5: Final C4A Model Creation and Capture Requirements

Require-ment ID	Req. description
REQ2.1	To ensure the realistic appearance of the C4A 3D model it should be created from captured 3D and video data of a real sign language translator (e.g. using a multi-view stereo capture setup).
REQ2.2	The C4A 3D model' s geometry-resolution should be sufficiently high to support realistic rendering, animations, well recognizable hand gestures and believable facial expressions/movements.
REQ2.3	To ensure believable facial expressions and emotions, the 3D model should be rendered with dynamic textures that are generated from real video footage.
REQ2.4	The captured sign language translator should not have a beard.
REQ2.5	The captured sign language translator should not wear glasses.
REQ2.6	The captured sign language translator should not wear black/very dark cloths.
REQ2.7	The captured sign language translator should have short hair or a tight haircut, which does not move when the head pose changes (e.g. no long wavy hair).
REQ2.8	In order to track the potentially fast movements of the captured sign language model the exposure time of all cameras must be configured in a way that motion blur is minimized.

4.2.2. Animation and Rendering

In order to achieve high expressivity, it is necessary to split the 3D model in distinctive animation components (i.e. face, body and hands), which can be controlled independently of each other. This can be based on the observation that, for example, the manifold of possible facial expressions does not depend on the displayed hand gesture. This means that the 3D model must be able to change its body pose, hand shape and facial expression. The following requirements are derived.

Table 6: Final C4A Animation and Rendering Requirements

Require-ment ID	**Req. description**
REQ2.9	The animation for face, hands and body should be independent of each other.
REQ2.10	The C4A 3D model should have a skeleton structure, which allows changing the body pose and showing different hand gestures.
REQ2.11	The C4A 3D model should have an animation structure that allows changing the facial expression.
REQ2.12	The C4A 3D model should be able to control the eye-gaze.
REQ2.13	The 3D Virtual Signer Language Translator should transform subtitles into animation parameters for the 3D model renderer.

4.3. Networking Requirements

The remote and heterogeneous architecture of the C4A project, being made up of different components each one having a different purpose and placed in different environments, implies a strong focus on the networking theme among them. The network is required to delivery management messages and multimedia contents. The latter in particular, require stringent constraints in terms of bandwidth in order to preserve the quality of the content needed for further processing.

For Networking requirements relative to the Pilot 1 Demonstrator (Phase 1) please refer to D2.2.

This section specifies the networking requirements for Pilot 2 Demonstrator (Phase 3) which covers the end-to-end content network monitoring and methodologies to fulfil them.

Table 7: Final C4A Networking Requirements

Require-ment ID	Description	Methodologies followed to fulfil the requirement
REQ3.1	3D Virtual Signer Language Translator should receive the subtitles from the Media Gateway thorough the Media Receiver.	The Media Receiver receives the subtitles from the Media Gateway and forwards them to 3D Virtual Signer Language Translator through TCP sockets.
REQ3.2	3D model renderer should receive the animation parameters for a particular signing sequence.	The 3D Virtual Signer Language Translator outputs the animation parameters as a JSON string, which is transmitted to the 3D model renderer as through a TCP socket
REQ3.3	Rendered 3D model images need to be provided to the mixer to overlay the 3D model onto the broadcast stream images.	The rendered images in uncompressed RGB format are provided to the mixer as an IP stream through TCP sockets
REQ3.4	The broadcast stream and the rendered images should be synchronized before overlaying.	The broadcast stream is decoded and potentially a fixed delay will be introduced to synchronize with the 3D model stream
REQ3.5	Rendered 3D model, should be overlaid at the correct position in the images of the broadcast stream.	The 3D Model will be located in the most appropriate position based on deaf community feedbacks
REQ3.6	The mixed frames need to be made available at the encoder to prepare media streams for MPEG-DASH adaptive streaming.	The mixed frames will be encoded by ffmpeg libraries using the H.264 code at different profiles and levels.
REQ3.7	Data packets that are transferred via network must contain timestamps in order to synchronize received packets in the processing modules.	

4.4. Encoding, Mixing and Streaming Requirements

The action of mixing the content coming from the 3D Model Renderer component and the main video coming directly from the broadcaster requires preliminary operations that must be taken into consideration, such as the synchronization of contents, uniformity to the same format, the position of the virtual signer into the whole video. Furthermore, before the content is sent, it must be duly converted to the appropriate format for shipping using the broadcaster infrastructure.

Requirements for encoding, mixing and streaming in the processing and rendering unit, combined with interactions from/to other C4A system components as well as their methodologies to fulfil them are listed in the following table.

Table 8: Final C4A Encoding, Mixing and Streaming Requirements

Require-ment ID	Description	Methodologies followed to fulfil the requirement
REQ4.1	The original video stream needs to be resized and mixed with the 3D model video stream into a single HD video stream.	Software mixing is performed with ffmpeg libraries.
REQ4.2	The single video stream should be encoded with different profiles to enable use of ABR protocols.	Software encoding is performed with ffmpeg encoding libraries. Hardware acceleration can be enabled using NVIDIA GPUs
REQ4.3	The different profiles stream should be packaged into ABR protocol supported by HbbTV and mobile devices.	The different streams are sent to the origin server which will package them into MPEG-DASH protocol, producing the proper manifest file
REQ4.4	The video is streamed to the client applications.	The client applications will download the manifest file from the origin server (or from CDN) and then will start downloading the video stream according to the available bandwidth

4.5. User Requirements

This chapter describes the derived requirements for the Content4All system from the Deaf community and other stakeholders. An important factor to system acceptance and thereby system success is the adaption of technical solutions to users' needs through user participation in the whole development process. Thus, in order to realize the Content4All solution, it is important for us to understand the expectations of the Deaf community. In WP2 *Specifications and requirements development,* HFC team followed a user-centred design-approach process, according to EN ISO 9241-210 and its conversion, in several steps of requirement analysis and their iterative evaluation methods (usefulness, usability, user experience). The analysis of all user and stakeholder requirements was conducted within T2.2 *Usability- and user-requirements and T2.3 Specifications for the user interface design*. It is based upon several pillars: data is collected using in-depth user-interviews with deaf and hearing signers, in-depth interviews with stakeholders (i.e. broadcasters, public service/ government, and public transport services), an online questionnaire for deaf and hearing signers, and a workshop with young deaf. Figure 5 shows the process of this in-depth analysis. After the data analysis, a list of the users' expectations towards Content4All was generated. The prioritization of these expectations according to the internal stakeholders (belonging to the fields of broadcast, innovation and research) was conducted in T2.3 via the Delphi method (Brady, 2015). Findings of the Delphi method established the final user requirements and specifications for the Content4All system.

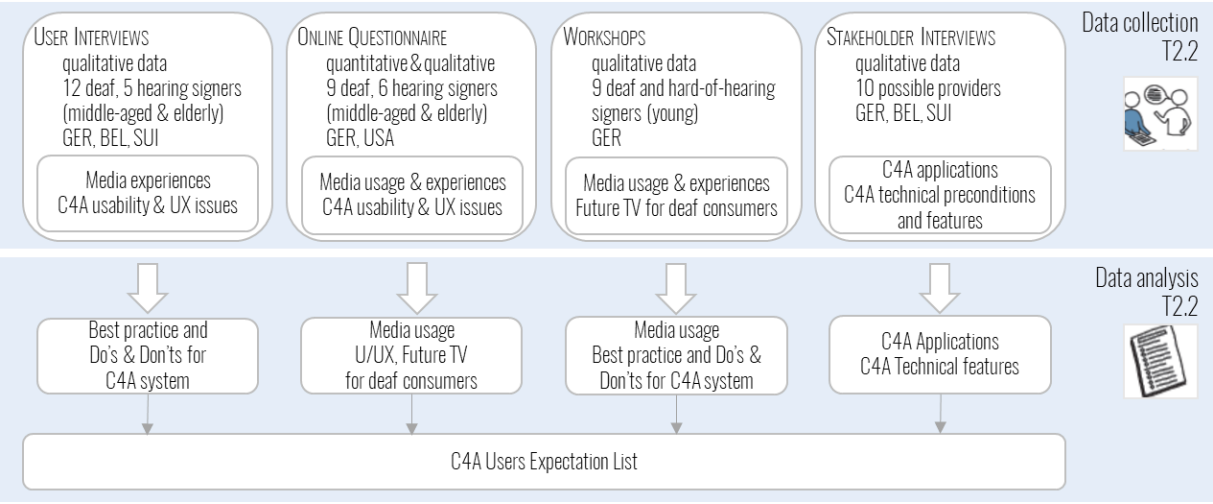

Figure 5: Process of T2.2 Requirement analysis

4.5.1. In-depth requirements analysis using user-centred design approach

In-depth interviews with users

Semi-standardized in-depth user interviews were conducted in 2018 with twelve deaf and five hearing signers from Belgium, Germany and Switzerland. Seven of the deaf interviewees were aged between 20 and 40 years, while five of them were older than 40 years and the average age of all five hearing signers was 33.4 years. The interview questions referred to the topics of demographic and general information, Deaf communication and sign language, current TV habits and media experiences, and expectations towards the CONTENT4ALL system. No hypotheses were assumed due to the qualitative and explorative method; all statements were analysed individually and reported in an aggregated way. The aim of the interviews was to find best practice examples as well as Do's and Don'ts for Content4All system.

Findings of the in-depth interviews with users. To understand user requirements at an early stage, the first most important aspect to understand is how the deaf user communicates (as television is a form of one-sided communication) with other deaf people and their hearing counterparts. Deaf people understand their surroundings only through visual input. Their mother tongue is sign language. Written language such as subtitles is equivalent to a foreign language for them. Findings show that among each other, deaf people communicate via sign language. If they are not facing each other, they use written communication ways like SMS, email or video chats. While communicating with other deaf people usually runs smoothly, most deaf are facing problems when communicating with hearing people. They rely on written and spoken communication forms as well as lip-reading (although a rare skillset) and interpreters.

The second part of the interviews dealt with experiences of the interviewees with media in general and TV in particular. This gave us insights into how deaf people currently use media, especially TV, and thus serve as best-practice and/or worst-case examples of what future Content4All system should contain or must avoid. Findings show that (1) the most used media device is the smartphone, followed by the PC. The deaf report difficulties with media in general as a lot of services (especially audio-visual media) are not accessible for the deaf due to an absence of sign language interpretation and/or subtitles. (2) All interviewees watch TV at least weekly. Traditional TV often does not offer subtitles, which is why many deaf switch to alternatives like Netflix, movies or media libraries. Besides a low availability of signed TV content, according to the asked, its comprehensibility strongly depends on the current signer. If he/she is not clearly understandable, e.g. because usage of wrong signs, the deaf has to fill this comprehensibility gap with increased cognitive workload which can be exhausting. The interviewees asked for a larger quantity of sign language interpreted content and subtitling of a high quality, especially also for e.g. entertaining programs and political discussion besides "typical" accessible shows like news. All such expectations including specific expectations towards Content4All in general, towards the virtual signer, and the user interface in particular were gathered and deduced into a users' expectations list.

In-depth interviews with stakeholders

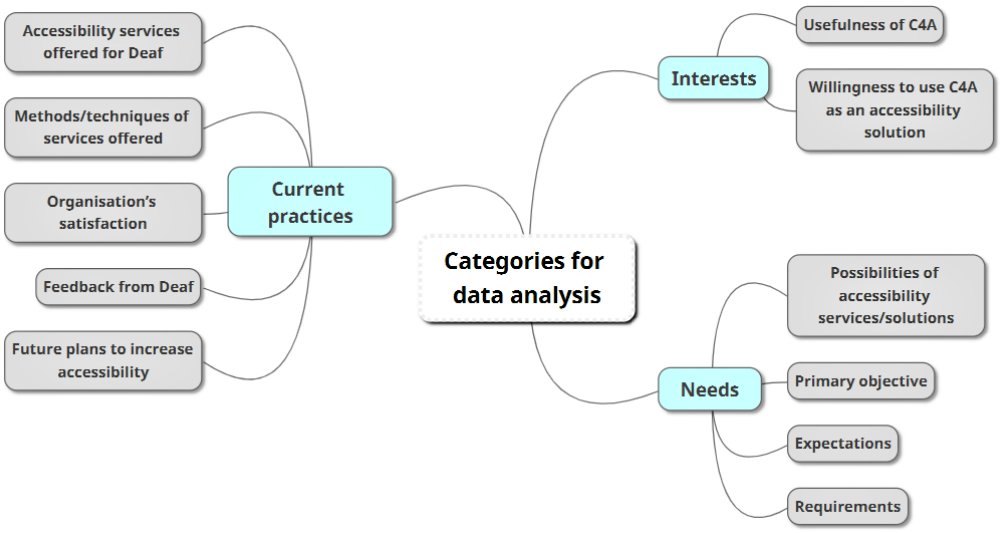

Figure 6: Categories and sub-categories of the semi-structured interview guide

For the analysis, industry-oriented interviews were conducted to collect generic ideas and expectations of the potential stakeholders (other than the deaf themselves) for a Content4All system (and services). Semi-structured interviews were used to gain insights into the current accessibility practices and expectations of the stakeholders. The interview questions were categorised into three major parts, displayed in Figure 6. In general, stakeholders are interested parties to a business or system; in this detailed case of Content4All project, they are for example European TV broadcasters, but also institutions with a public mandate as well as businesses interested in using the Content4All system commercially. A total of ten different organisations were interviewed: Four interviewees belong to the media broadcast industry, two located in Germany and one in Belgium and Switzerland respectively, two interviewees belonged to the government sector, both located in Germany, and three interviewees are public transport services (railway) all located in Germany[1]. Its results inform business models and dissemination activities carried out in WP6.

Current accessibility practices and feedback from deaf community. It was important to gain an overview of the extent to which the acquired stakeholders provide accessibility to people with disabilities, especially the deaf. All broadcaster stakeholders offer multiple services to people with disabilities. All four stakeholders in this category primarily provide accessibility in the form of subtitling and a few programs in sign language. Two out of three government stakeholders have an obligation to provide live sign-language translation and live subtitling during government sessions. They hire external service providers to provide live subtitling and live translation for TV transmission of parliament debates into sign language as well as to manufacture information videos (in sign language) on their websites. As compared to the broadcaster and government stakeholders, the public transport stakeholders provide a relatively small amount of accessibility services that are specially designed for the deaf, due to economic reasons.

Most interviewed stakeholders (all broadcasters, two of the government stakeholders and one public transport stakeholder) are satisfied with their current accessibility practices as many receive positive feedback, but to a large extent, they also receive negative responses from the deaf regarding their current accessibility practices. Common complaints from the deaf community are, for example, the dissatisfaction with the limited number of shows offered in sign language/with subtitles or

[1] An attempt was made to contact organisations from other industries as well (companies offering online services, events etc., virtual reality and gaming companies), but their lack of any accessibility services for the deaf resulted in non-participation in an interview.

that some content is solely available online. A closer exchange with the deaf community would enable stakeholders to better meet the demands of the target group.

Stakeholder needs and expectations. Based on the current problems and feedback from the Deaf community, most of the interviewed stakeholders are working on future plans to increase accessibility. Some stakeholders have shown high interest in using the Content4All system for possible accessibility services. After establishing each stakeholder's interest and willingness to use the Content4All system, they were further interviewed to ascertain the services that they would like to implement in their respective operations/business areas using the Content4All technology and what they expect from the Content4All system. All stakeholders emphasized on user acceptance and cost-effectiveness of Content4All. Others expressed their interests in good user-friendliness and Content4All's application on multiple devices. Technical expectations such as inclusion of parallel signals to enable switching on/off the sign-translated content and time synchronicity of the interpretation and video content were expressed. This helped us derive their needs and expectations. All expectations were rephrased and listed in the user expectations list.

Workshop

In order to understand expectations from all age groups, a workshop was conducted with young deaf and hard of hearing people. Nine students from a school with a special educational profile for "Hearing and Language" located in Germany participated in the workshop (all students were deaf or hard of hearing, age ranging from 12 to 18 years). The school selected the participating students upon two conditions: their mother tongue had to be sign language and they had to be familiar with the use of different types of media. The workshop's focus was to gain insights into the media usage of young deaf and hard of hearing people and the aspects that are crucial for them to use certain media. It is very important to ask these young people as a sub-group of deaf as we suspect that young deaf will make up a large part of Content4All's future target group. For this, a combination of different tasks was used:

- Usage of devices and media services: The students identified their weekly usage of different devices and media services.
- Benefits and shortcomings of media services functions: The students evaluated different functions of media and their services.
- Accessibility of audio-visual media: The students compared benefits and shortcomings of different videos that are tailored to the deaf.
- Drawing activity: The students developed individual interfaces for a possible new accessible format.

Findings of the workshop. In the first two tasks, we found that the smartphone was the most used media device among the students followed by TV. All of them denounce the lack of accessible TV programs that interests students. On the one hand, accessible content for example in sign translated language is not interesting to the students whereas on the other hand, content that students want to watch shown by private broadcasters often does not offer accessible sign language translation or even subtitles. Thus, they resort to alternative media providers like YouTube, Netflix or AmazonPrime. In the third task of the workshop, clips of five different videos were shown to the students and were compared to each other. The videos varied regarding their availability for deaf (subtitles, signing) and their genre of the content. It was found that the attractiveness of the content is important for the choice to watch a certain show, but other aspects such as social identification influence the choice for following audio-visual media.

Expectations of young deaf. In the last task, the students developed their own ideas on how their personal interface of a new accessible audio-visual media format would look like. Each student designed at least one draft, primarily for TV/computer and tablet. 23 sketches on how a system should (ideally) look, contributed to deriving expectations for the Content4All system. These were added to the user expectations list.

Online questionnaire

Quantitative and qualitative data was collected using an online questionnaire (Figure 7) to identify the current usage of media devices and media services by the deaf, the problems they face concerning the same and their likes, dislikes and preferences of the current media services. Moreover, specific questions were asked to understand what aspects of a future TV for deaf consumers are important, especially features (and other characteristics) of a virtual signer. The responses were then translated into user expectations. The questionnaire was filled out by nine deaf or hard of hearing persons and 6 sign language interpreters from Germany and the USA. All participants can communicate in at least one sign language and one phonetic (spoken) language. Hearing participants were required to respond several questions from a deaf person's point of view i.e. their opinion on deaf people's expectations from novel TV technologies based on their knowledge and experience. The questionnaire was built in two versions: English/British Sign Language (BSL) and German/German Sign Language (DGS). The questions asked concerned demographic information of the participants, deaf people's usage of media devices and media services, and expectations from novel television technologies.

Media usage. The most often used device is the smartphone, followed by PC and TV. Even though the deaf use traditional TV less frequently, it was found that they do like to watch TV. In terms of communication, findings show that WhatsApp is the most popularly used media service, followed by Facebook, Facebook messenger and Skype. These media services allow hearing and deaf/ hard of hearing people to communicate with each other, especially thanks to texting, video calling and sending pictures. Some functions such as voice calling and voice messages are however useless for the deaf. In terms of 'TV habit', YouTube is used at least multiple times a week, Netflix is used less commonly followed by Amazon Prime. A comparison between YouTube, Netflix and Amazon Prime showed us that YouTube is the only media service used by the deaf to watch videos/channels in sign language, for example, YouTube channels by deaf influencers. To watch movies, series etc. the participants prefer (apart from YouTube) Netflix as it offers a large library and high-quality subtitles in different languages. YouTube and Amazon Prime on the other hand, offer poor or false subtitles if any. Based on these findings, user expectations for Content4All system were deduced.

Expectations from novel TV. It was found that half of the signers prefer to watch TV in both sign language and with subtitles; reasons are for example to allow choosing one of the options depending on the kind of show or to be able to judge the signed content better by comparing it to the subtitles. In order to get insights on the best practices and worst cases, all participants were asked about their likes and dislikes of current sign-translated TV shows including the likes and dislikes of the interpreter and what improvements can be made to them. According to the participants, sign language should be considered an integrated part of the production process and not an afterthought. Consideration should be made to ensure quality of the sign language and of how it is displayed, for e.g. interpreters should be made more visible (positioned larger) to clearly understand the signing.

Expectations to the C4A virtual signer and TV interface. As sign language is a visual language, the movement and positions of the virtual signer's physical features is important to understand. This information is taken into consideration for modelling the Content4All virtual signer. Participants were explicitly asked which features a virtual signer i.e. Content4All virtual signer should have. According to the participants, the most important aspects are signs and gestures and its hand and finger movements and positions, followed by body movements and position, and eye movements and position. Gender and age of the virtual signer are considered not relevant by most of the participants, although some suggested for these characteristics to be adapted to the viewer or the target group of the show. Most participants specified that the clothes of the virtual signer should be dark, plain (nothing distracting) and uni-coloured. There should be contrasts between skin colour and clothing as well as between the clothing and the background. More than half of the participants conveyed that they prefer the signer separated from the TV picture. All these inputs were utilised to derive user expectations for the Content4All system and were summarized into the user expectations list.

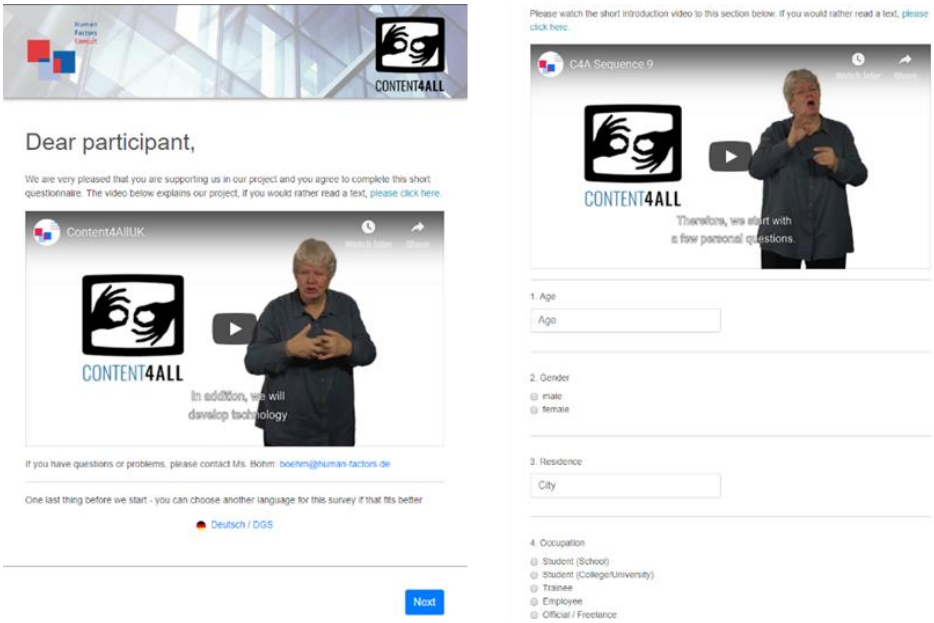

Figure 7: Screenshots of the online questionnaire

4.5.2. Prioritization of user expectations and final user requirements and specifications for Content4All system

Through the multi-method analysis (data from the users and external stakeholders collected through the user expectations analysis in T2.2), user expectations for Content4All system were derived and summarized in the form of an expectations list. All expectations were then categorised into four classifications: General expectations, technical/usability expectations, C4A TV interface & virtual signer expectations and C4A accessibility features expectations. All repetitions were removed and the expectations were paraphrased when required.

In order to derive the final user requirements and specifications, a prioritization process was conducted with help of the Delphi method. The aim in using this method was to determine which aspects should be included in the Content4All system (REQ) and which aspects should be excluded i.e. are out of scope (OoS) for the Content4All project. As per the Delphi method, a total of three rounds were conducted in order to determine the user requirements for the Content4All system. Ten representatives from five internal stakeholders (Content4All project partners) belonging to the fields of broadcast, innovation and research participated in the method. Figure 8 illustrates the overall process.

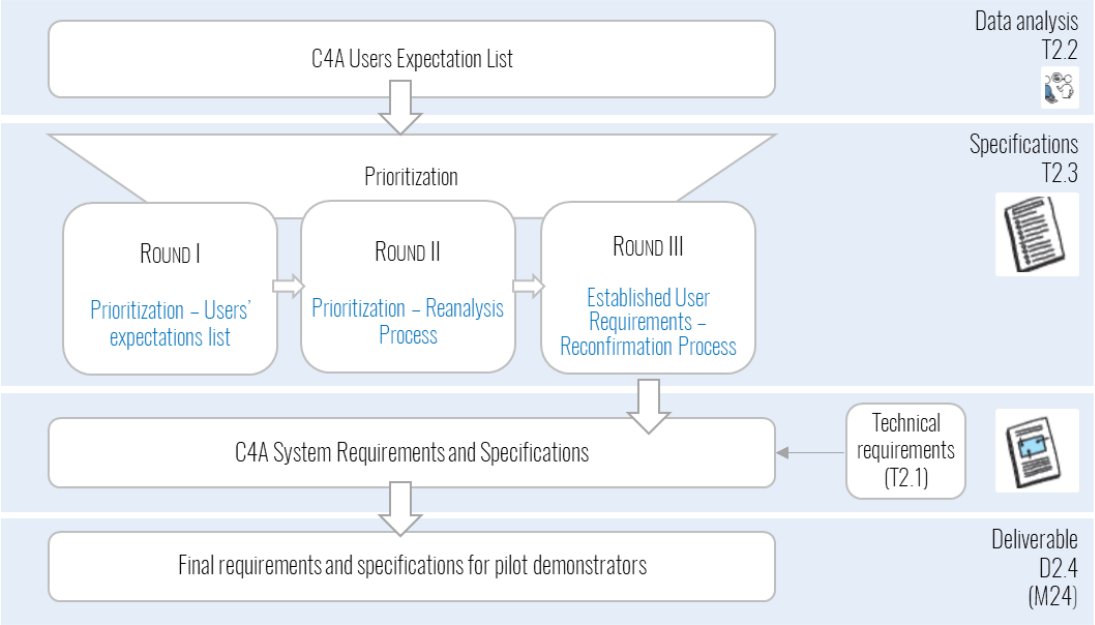

Figure 8: Prioritization process - Delphi analysis

Prioritization of user expectations list (Round I)

Within the first round, the user expectations' list was handed out to the 10 representatives of the different internal stakeholders. They were asked to give their prioritization for each user expectation based on their expertise contributing to the project / CONTENT4ALL system to be developed. For this, they were provided with four 'prioritization levels' to rate each expectation:

1. MUST (M): MUST requirements are minimum requirements of a product/service that are absolutely essential and necessary to meet the business needs and thus for products'/service's success. They are the Minimum Usable Subset requirements.
2. SHOULD (S): Meeting SHOULD requirements is also important for products/services success but not that essential as MUST requirements are. SHOULD requirements are less time-critical and should be met if they do not interfere with MUST requirements.
3. COULD (C): COULD requirements are less critical for products/services success. Often, they are useful extensions for the product/service that can be implemented if all MUST and SHOULD requirements have been met.
4. WON'T (W): WON'T requirements are those least-critical, not appropriate or with lowest payback. For momentary product/service development they are not relevant but can be kept in mind as useful ideas for future product extensions.

All participants gave each expectation the rating of 'Must, Should, Could or Won't'. The ratings were then summarized into three tables. The first table listed all expectations for which the analysis revealed a clear overall tendency to include as a user requirement in the CONTENT4ALL project, i.e. a large majority of the participants (over 60%) stated a Must or Should. If the median and mode differed from each other, a subcategory was defined (M/S). The second table listed all expectations for which the analysis revealed a clear overall tendency to not include (OoS) as a user requirement in the project, i.e. a large majority of the participants (over 60%) stated a Could or Won't. Again, if the median and mode differed from each other, a subcategory was defined (C/W). The third table listed all expectations for which the ratings revealed no clear overall tendency to include or exclude (no large majority for Must & Should or Could & Won't was found). This was indicated with a new category S/C.

Reanalysis of prioritization (Round II)

To finalize the decisions resulting from the participants' ratings in round 1, they were asked to indicate again if they agreed (A) or disagreed (D) with the decisions listed in the first two tables (expectations that should be included and should be excluded). For the third table where no clear decision could be reached, they were asked to provide a final statement, if the expectation should be included as a requirement (Y) or excluded (N). Additionally, they had to explain their decision. Each list was analysed separately. Similar to round 1, a rating of the majority of the participants (above 60%) determined if the expectations would be included or excluded, thus regarding them as established requirements (REQ) for the Content4All system or out of scope (OoS) for the Content4All system. As there were still disagreements between the participants (those features whose ratings were 50%-50%) regarding few features, the prioritization was decided in favour of the technical partners.

Established user requirements – reconfirmation process (Round III)

A third and final round was required to be conducted in order to finalize on the list of established requirements. This round enabled fine-tuning of the expectations and ultimately final consensus on those expectations that should be included and those that should not, giving us the established user requirements and specifications for the C4A system. Based on the analysis of round 3 of the Delphi method and input from the technical requirements, the final user requirements and specifications for Content4All system were established. The final user requirements and specifications are listed in Table 9.

Table 9: Final user requirements and specifications for Content4All system

Topic	ID	Description
General	REQ5.1	Users accept C4A system and the virtual signer.
	REQ5.2	C4A system provides broadcasted programs accessible for the Deaf.
	REQ5.3	C4A system is user friendly.
	REQ5.4	C4A system is available next to the TV for other devices, especially for web-enabled devices like smartphone, tablet and PC.
	REQ5.5	C4A system does not contain any advertisement.
Technical / Usability	REQ5.6	C4A system works without technical problems.
	REQ5.7	C4A system is time efficient and cost-effective, can be used with affordable material, does not require high-end devices.
	REQ5.8	C4A system audio-visual content is provided in high resolution.
	REQ5.9	An acceptable time synchronicity with respect to the actual video, audio, and video with sign-translated content for both pre-produced as well as live programs is ensured.
	REQ5.10	C4A system provides a function where the user can filter the offered audio-visual media content by different options (i.e. the availability of sign language).
C4A TV Interface & Virtual Signer	REQ5.11	C4A virtual signer has a pleasant appearance.
	REQ5.12	C4A virtual signer's arm movements are visible.
	REQ5.13	C4A virtual signer wears long plain and neutral sleeves (no strips, checks or any texture) in contrast to skin colour.
	REQ5.14	C4A virtual signer may not wear disturbing/distracting elements like long earrings, tattoos, glasses, beard or moustache, nail polish, tattoos.
	REQ5.15	C4A virtual signer looks realistic (in contrast to an animated puppet).
	REQ5.16	C4A virtual signer presents the following features in very high quality (clear and precisely perceivable): signs, gestures, hand movements and –position, finger movements and –position, body movements and –position, facial expression/mimics and emotion primarily mouth, cheeks, eyes and eyebrows.
	REQ5.17	C4A virtual signer's signs fit to its facial expression.
C4A Accessibility Features	REQ5.18	C4A system provides sign language interpretation for informative TV content (where the focus lies on the transported information, i.e. news, documentation
	REQ5.19	C4A system provides audio-visual content with sign language interpretation for a specific domain scenario, e.g. weather or news.

Topic	ID	Description
	REQ5.20	C4A system provides sign language interpretation with high quality, i.e.: • the respective national sign language (no word to word translations from spoken language) • with correct vocabulary: standardized vocabulary, gestures/signs known to the community • with correct grammar • natural facial expression/mimics and gestures: fluent, not exuberant/extreme/exaggerated • transportation of atmosphere and emotions of signed context by using correct facial expression/mimics • in reasonable signing speed (not to fast, not to slow)
	REQ5.21	C4A system user can switch the sign language interpretation on or off.

A summary of established user requirements is displayed in Figure 9. Even though all expectations collected in T2.2 are regarded as important for accessible TV-content for the deaf, some of them are just out of scope to realize them all in the current project. Thus, these 'out of scope' requirements can still serve as an orientation for future projects. Especially the high demand for subtitles, which cannot be fulfilled in this project, should be kept in mind.

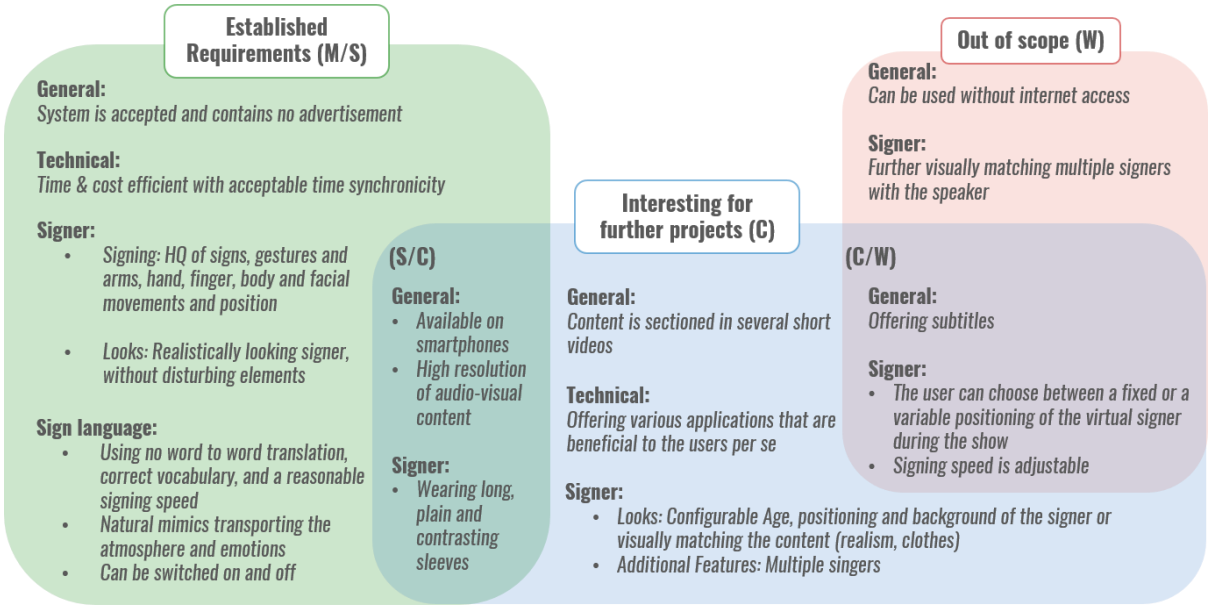

Figure 9: Summary of the established user requirements and out of scope requirements

4.6. Evaluation Tools

The CONTENT4ALL project follows a user-centred design approach, i.e. users are involved during requirements collection, prototyping and testing. This section focuses on providing information about evaluation concepts, tools and methodologies for the planning of the user tests in WP5 as well as deduced requirements for the evaluation phase.

Usability is defined as the "extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (EN ISO 9241-11). General classes of usability measures address the key criteria effectiveness, efficiency and satisfaction. Specifications of these classes however vary with respect to the product, system or service (Brooke, 1996). Usability is one key aspect in describing user experience (UX), along with emotional and aesthetic aspects. UX comprises the entire set of affects that result from the interaction between a user and a product. This includes sensory stimulation (aesthetic experience), the meaning, which is attributed to the product (experience of meaning experience) and the feelings and emotions that emerge during the interaction (emotional experience, Hekkert 2006). The relationship between usability and UX is intertwined (Vermeeren et al.,2010). One important difference is that usability focuses on the task performance (execution time) whereas UX components additionally take into account the actually perceived experience (user's affects, motivation and expectations). In comparison to usability, UX thus is holistic in nature. UX is often classified as "subjective" and usability as "objective". However not all components of usability are objective such as the "satisfaction" component which is seen as a part of UX. Furthermore, studies rather focus on the perceived usability of a system or product for which questionnaires partly based on ISO 9241-110 (Dialogue principles) are available (e.g. Isometrics, Isonorm 9241/10, System Usability Scale).

The UX of a system consists of at least two qualities: pragmatic or instrumental qualities and hedonic or non-instrumental qualities (Hassenzahl, 2008; Hassenzahl and Tractinsky, 2006; Thüring and Mahlke, 2007; Schrepp, Hinderks, & Thomaschewski, 2014). Pragmatic or instrumental qualities refer to product qualities such as its usefulness and usability. Hedonic or non-instrumental qualities focus on aspects such as the product's aesthetics and its haptic quality, i.e. the feel of the product. The user's feeling of so called "be-goals" (e.g. "being competent", "being related to others", "being special") and general needs (e.g. for novelty and change, personal growth, self-expression and/or relatedness) are addressed by these qualities (Hassenzahl, 2008, p. 2). The above-mentioned components result in the user's overall appraisal of the product or system which in turn characterises their opinion of the product, attitude towards the product, behaviour towards the product, such as the decision to use the product or not, if yes then how often, intentions towards the product, such as probability of migrating to other similar products, etc.

Aspects such as quality of service (QoS), quality of experience (QoE) and UX are closely linked with each other. QoS refers to the technical characteristics that are related to the overall technical performance of a system. On the other hand, QoE refers to the user's perception of the performance of a system or service (Dieplod, 2012) and UX helps in exploring how users feel about using a product, system or service (Vermeeren et al., 2010). QoE and UX focus on the perception of the user and thus built a base for user evaluation requirements and planning. Based on the key points mentioned above, the requirements REQ6.1 – REQ6.6 for C4A evaluation are compiled and listed in Table 10.

During the evaluation phases, a variety of methods could be used, with the exact selection always depending on the issues in question and/or the respective development goals. The approach should be geared to a user-centred design process according to EN ISO 9241-210 and its conversion in several steps of requirement analysis and their iterative evaluation methods. Depending on the evaluation stage, different qualitative and/or quantitative methods are used. The information acquired is then evaluated and processed.

The goal of the development is to incorporate scales and measurements relevant to the UX of C4A a way that is as faithful as possible to the empirically tested sources while adhering to the requirements of efficiency and validity. For this purpose, an encyclopaedia of measurement methods was compiled. The encyclopaedia has different categories, according to which the methods are classified. These include classification into the type of UX aspect addressed (pragmatic, hedonic, social), the type of collected data, the study type, collection method, time of evaluation as well as the methods' strengths and weaknesses, and their costs. The categories are described in more detail in D5.3. With regards to these categories all of the methods within the

encyclopaedia were evaluated within the HFC research group in terms of their relevancy for CONTENT4ALL demonstrator evaluations. Twelve out of 44 methods have been selected as potentially serving the purpose of testing components (formative testing) or demonstrator 2 (summative testing) of the CONTENT4ALL system. With regard to the user tests, we thus propose the requirements RE6.7 – REQ6.13 for C4A system evaluation tools (see Table 10). These requirements mainly relate to the evaluation process and evaluation methods used within the project. With the completion of the encyclopaedia, these requirements have already been addressed and in terms of method selection, they have been fulfilled.

Detailed information of user test plans for evaluation can be found in D5.3.

Table 10: C4A Evaluation Tools Requirements

Require-ment ID	Req. description
REQ6.1	Test plans should focus on the quality of service (technical aspects), quality of experience (user's perception of performance) and user experience (user's impression) of C4A system.
REQ6.2	Methods should be identified to conduct user tests for the overall UX of C4A system (summative evaluation).
REQ6.3	Definite test periods should be established right from before development of C4A system to after completion of the development of C4A.
REQ6.4	User test plans should be created for every test period i.e. before development of C4A system, during development of C4A system and after development of C4A system.
REQ6.5	Various QoS, QoE and UX components need to be investigated and relevant components need to be identified for C4A. Focus should be given to UX components.
REQ6.6	The UX components should include pragmatic qualities and hedonic qualities.
REQ6.7	Data collection tools should be selected in order to collect data relevant to the project at all stages of UX evaluation.
REQ6.8	Various evaluation methods have to be identified for assessing the selected UX components (formative evaluation).
REQ6.9	Objective and subjective evaluation tools have to be selected with respect to the chosen UX components for C4A.
REQ6.10	Evaluation tools in the form of questionnaires should be developed that suite the UX requirements for C4A.
REQ6.11	Individual items should be identified with respect to UX requirements that can be incorporated in the developed questionnaires.
REQ6.12	Evaluation tools should be developed or existing questionnaires should be modified to fit the needs of C4A.
REQ6.13	The selected evaluation tools (subjective or objective) should be able to assess each component individually so that each component can be individually interpreted.

4.7. Business models

Different business models for the C4A system Pilot 1 Demonstrator (Phase 1) are defined: for further information please see D2.2.

For what concerns the Pilot 2 Demonstrator (Phase 3), since it is a proof-of-concept, Business models will be analysed at a more advanced exploitation stage.

4.8. Requirements and Matching C4A objectives

For the Pilot 1 Demonstrator (Phase 1), the matching between Requirements and C4A objectives remains unchanged. For further details please see D2.2.

As concerns the Pilot 2 Demonstrator (Phase 3), in this section, the identified requirements are matched to C4A main objectives and listed in the following table.

Table 11: Final C4A Requirements and Matching to Project Objectives

Objectives	Requirements
1. Translation from subtitles to sign-language 3D poses, 3D model generation and animation	1.2, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 2.10, 2.11, 2.12, 2.13
2. Signal processing	1.1, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 4.1, 4.2, 4.3, 4.4
3. Experimental performance evaluation	6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, 6.10, 6.11, 6.12, 6.13
4. Building a proof-of-concept demonstrator	5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 5.10, 5.11, 5.12, 5.13, 5.14, 5.15, 5.16, 5.17, 5.18, 5.19, 5.20, 5.21

5. Conclusion

This document described the use cases and final technical, user and evaluation requirements of C4A system Pilot 1 (Phase 1) and Pilot 2 (Phase 3) Demonstrators as well as their matching to projects objectives, based upon the initial system and user requirements and specifications for Pilot 1 (Phase 1) and Pilot 2 (Phase 3) Demonstrators which are addressed in "D2.2 Initial requirements and specifications for pilot demonstrators".

Both, use cases and final requirements for C4A system that were derived based upon the main system components, were identified within this document and mitigation solutions were provided. The use cases and requirements refer to the Pilot 1 Demonstrator (Phase 1), where a sign-interpreter will be captured and inserted into existing TV during the defined reference scenario of News, and Pilot 2 Demonstrator (Phase 3), where the development of an automatic sign-interpretation technology for the scenario of News is proposed.

Additionally, the potential of using the Phase 2 models for animating a photo-realistic human virtual signer is being explored.

Further deliverables covering the details of the architecture, reference scenario, final use cases and requirements as well as planning of C4A demonstrator planning are listed below:

- D2.1 "Initial Reference System Architecture"
- D2.4 "Final Requirements and Specifications for Pilot Demonstrators"
- D5.1 "Pilot 1 Demonstrator Architecture, Integration Plan, and Evaluation Methodologies"
- D5.5 "Integrated Pilot 1System and Evaluation Report"
- D6.3 "Business Model Analysis"

6. References

Bargas-Avila, J.A., & Hornbæk, K. (2011). Old Wine in New Bottles or Novel Challenges? A Critical Analysis of Empirical Studies of User Experience, Proceedings of CHI 2011 (May 7–12, 2011), Vancouver, BC, Canada.

Brady, S. R. (2015). Utilizing and adapting the Delphi method for use in qualitative research. *International Journal of Qualitative Methods*, *14* (5), doi: 1609406915621381

Bradley, M.M. & Lang, P.J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. Journal of behaviour therapy and experimental psychiatry, 25(1), 49- 59.

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, *189*(194), 4-7.

Dieplod, K. (2012). The Quest for a Definition of Quality of Experience. Qualinet Newslet, 2, 2-8.

EN ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability

EN ISO 9241-210: Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems

Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In J. Ziegler & G. Szwillus (Eds.), Mensch & Computer 2003. Interaktion in Bewegung, pp. 187-196. Stuttgart, Leipzig: B.G. Teubner.

Hekkert, P. (2006). Design aesthetics: Principles of pleasure in product design. Psychology Science, 48(2), 157-172.

Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. Holzinger, A. (Ed.): USAB 2008, LNCS 5298, pp. 63-76.

Schrepp, M., Hinderks, A., & Thomaschewski, J. (2014). Applying the user experience questionnaire (UEQ) in different evaluation scenarios. In *International Conference of Design, User Experience, and Usability* (pp. 383-392). Springer, Cham.

Vermeeren, A. P., Law, E. L. C., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010, October). User experience evaluation methods: current state and development needs. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries* (pp. 521-530). ACM.

Partner Short Names

Short Name	Name
FIN	Fincons Group AG
UNIS	University of Surrey
HHI	Fraunhofer Institute for Telecommunications Heinrich Hertz Institute
HFC	HFC Human-Factors-Consult GmbH
TXT	Swiss TXT AG
VRT	Vlaamse Radio -en Televisieomroeporganisatie