Initial Requirements and Specifications for Pilot Demonstrator

Deliverable D2.2

CONTENT4ALL File ID:	C4A D2.2 Initial Requirements and Specifications for Pilot Demonstrator.docx
Version:	1.1
Deliverable number:	D2.2
Authors:	Sandra Böhm (HFC)
Contributors:	Wolfgang Paier (HHI), Pasquale Panuccio (FIN), Thanuja Mallikarachchi (UNIS), Robin Ribback (TXT), Vaishnavi Upadrasta (HFC)
Internal reviewers:	UNIS, FINCONS
Work Package:	WP2
Task:	T2.1, T2.2, T2.3
Nature:	R – Report
Dissemination:	PU – Public
Status:	Final
Delivery date:	21.06.2018

Version and controls:

Version	Date	Reason for change	Editor
0.1	19/01/2018	First draft	HFC
0.2.1	06/02/2018	Contributions from HHI	HFC
0.2.2	16/02/2018	Contribution from UNIS	HFC
0.2.3	22/02/2018	Contribution from HHI	HFC
0.2.4	22/02/2018	Contribution from FIN	HFC
0.3	23/02/2018	Finalize first version	HFC
0.4.1	26/02/2018	First review comments addressed	HFC
0.4.2	28/02/2018	Contribution from HHI, UNIS & FIN	HFC
0.4.3	01/03/2018	Adaptations	HFC
0.4.4	04/03/2018	Adding use cases and requirements for business models	HFC
0.5	05/03/2018	Finalize second version	HFC
0.6.1	06/03/2018	Review comments addressed	HFC
0.6.2	07/03/2018	Review 3.8 & 4.7	STXT
1.0	07/03/2018	Finalize for Submission	HFC
1.1	21/06/2018	Reviewers comments addressed & Finalization for Resubmission	UNIS, HFC

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.

Disclaimer: This publication reflects only the author' s view and the European Commission is not responsible for any use that may be made of the information it contains.

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, these deliverable about initial use cases and requirements focuses the CONTENT4ALL system phase 1 demonstrator in M12. In 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.

The purpose of this document is to report use cases and initial requirements, which cover the four CONTENT4ALL system components *Broadcaster*, *Remote Studio*, *Processing and Rendering Unit* and *User Terminal* and address the objectives of sign-interpretation extractions, signal processing, creating business models and building a large-scale demonstrator. The defined use cases and requirements describe the interaction of the components as well as the characteristics the system should demonstrate at M12. The use cases address the recording of the remote studio data, their transmission to and within the processing and rendering unit, estimation of the handshape, the rendering of the 3D model of the sign interpreter, the content preparation in the processing and rendering unit, the transmission of the 3D model rendered video stream and main broadcast streams to the user terminal and business models. The requirements refer to the content generation, the rendering/model generation, the networking, the encoding, mixing and streaming, the users the evaluation tools and business models. All use cases and requirements are matched to the project objectives.

Table Of Content

List of Figures

List of Tables

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 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 phase 2 an automatic sign-interpretation technology for deaf TV watchers will be developed, which is displayed via an animated photo-realistic human signer, and exemplary demonstrated in a defined application scenario.

This document provides the use cases and requirements which are defined for CONTENT4ALL system for phase 1 demonstrator at M12. They address the interaction of the components as well as the characteristics the system should contain at M12.

1.2. Structure of the Document

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

In chapter 4, initial requirements derived for CONTENT4ALL application are reported. As shown in the use cases in section 3, the requirements also cover broadly the four main C4A system components *Broadcaster*, *Remote Studio*, *Processing and Rendering Unit* and *User Terminal* as well 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 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. Learning-based sign classification and segmentation mechanism will be developed.
- Phase 2 proposes to develop an automatic sign-interpretation technology for the scenario of News (see section 2.3) and explore the potential of using the phase 1 models for animating a photo-realistic human signer.

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

2.1. System Architecture and Component Description

A high-level representation of the C4A system architecture shows the logical composition and the interaction among the system's 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.

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 skeleton data along with the video is streamed to the processing & rendering unit component.
3. The processing and rendering unit processes input data and reproduce the sign language through a 3D photorealistic model.
4. Both video streams (original and mixed one) are sent to the users' terminals.

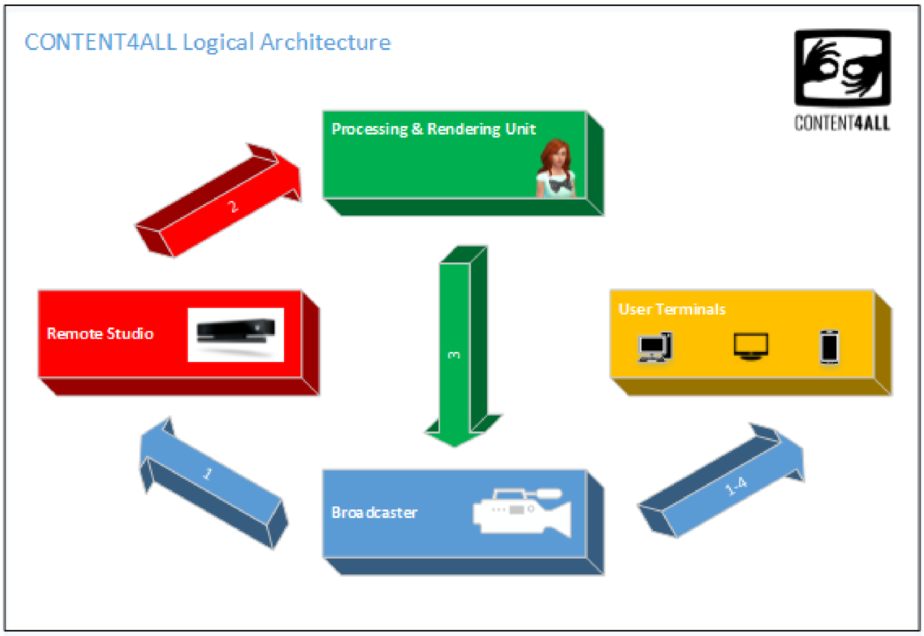

Figure 1: C4A Overall Logical Architecture Phase1

Figure 2 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 is replaced by the remote studio where a sign interpreter signs the video content.
- 3D rendering data are captured and transmitted to the photorealistic 3D model renderer and transformed to a photorealistic 3D human signer.

The workflow can be divided in four components:

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

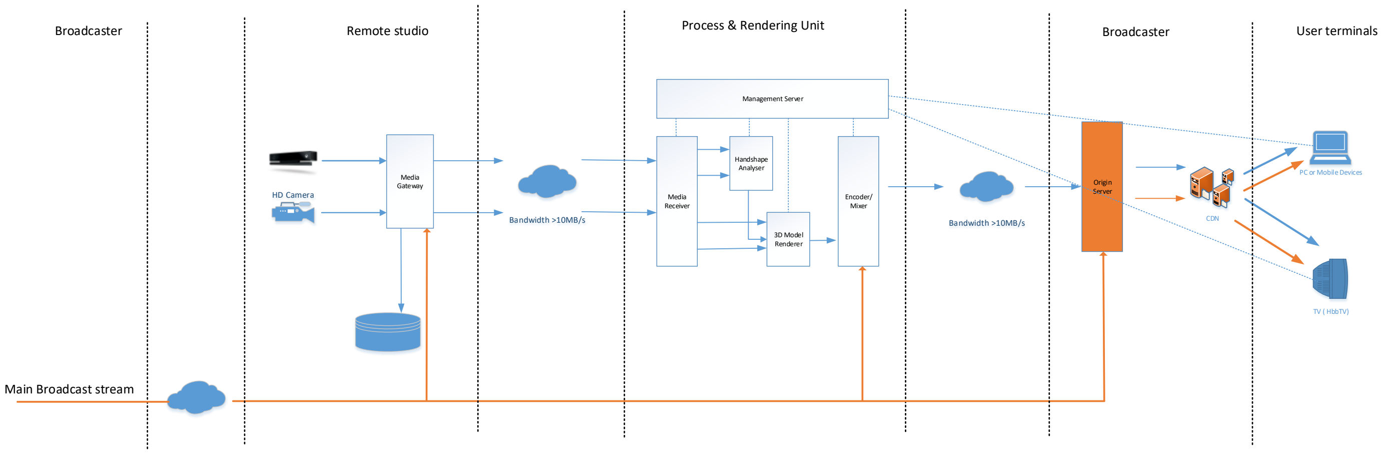

Figure 2: C4A Overall Architecture Workflow

A detailed description of C4A system architecture and its components can be found in D2.1.

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 (SWISS-TXT) 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 two phases of the project. In the first phase, all three TV shows will be produced and translated using the remote studio, while the second phase will focus only on weather news to demonstrate the results of the automated sign translation process.

A detailed description of the reference scenario, and differences between the two phases can be found in D2.1.

2.3. Phase 1 Demonstrator

During the project two end-to-end demonstrations are planned, in order to demonstrate the technologies developed during the C4A project. The first demonstration will be ready at M12 of the project while the second demonstration will be available at M33. As this deliverable focuses only on phase 1 demonstrator, the following is true:

C4A will set up two similar environments at the project's partner studios (SWISS-TXT and VRT) in order to run and test the developed framework. The sign interpreter is located at the remote studio and will start signing in real time while his movements are recorded by a Kinect camera. In addition, facial expressions are recorded by the studio camera. Such information will be 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 will translate the input to a stream of metadata representing the intelligible format of handshapes for rendering. The metadata stream will be the input to the photorealistic 3D model renderer along with the video, skeleton and depth streams data provided by the sensors (aka Kinect and camera) placed in the remote studio. The model renderer component will render a 3D model. The 3D photorealistic renderer component will add the rendering of the handshape. The model's stream will be 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 will be further adapted (i.e. encoded in different formats) and delivered to HbbTV and mobile devices via broadband.

The mixed stream will be then sent to users through the origin server component. The translated content will be streamed to different clients via the broadband connection. This content will be 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 will be 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.

During the phase 1 of the project, all data generated from remote studio will be collected and stored into the storage systems already available in the broadcaster's infrastructure. This data is needed in order to create a model for the translator and for the new 3D model that will be rendered during the phase 2 of the project.

Details of the phase 1 and phase 2 demonstration will be included in D5.1 and D5.2, respectively.

3. Use Cases

As already mentioned, the C4A system is built up of the four main functional components *broadcaster, remote studio*, *processing and rendering unit* and *user terminals*. All generate special content which is essential for the

- interactive content generation actions for the reference scenario, defined in D2.1, and recapped in section 2.2 above,
- the phase 1 demonstrator (M12), described in D2.1, and
- the project's overall objectives (see D2.1).

This section defines the main use cases which illustrate the interaction of the C4A system components for phase 1 demonstrator at M12 in terms of the project's objectives. All use cases will be described with preconditions, postconditions and at least one main success scenario. Preconditions describe the states of the C4A system that have to exist before the use case can start. Postconditions deal with the states of the system after the use case has been successfully completed. The main success scenario describes what the phase 1 demonstrator will (or will not) achieve, as a result of the respective use case.

The presented use cases include all C4A system components: Sections 3.1 until 3.6 address the remote studio and the processing and rendering unit as well as their interaction, section 3.7 and 3.8 list use cases for the broadcaster and the user's terminals. The last two sections report on the link between use cases and objectives, and provide a risk assessment: 3.9 matches the use cases to the C4A system objectives, and 3.10 performs an analysis of potential risks of the use cases. Thereby, main functionalities of C4A system are covered and linked to the actors (mainly system components) that are addressed.

3.1. Recording Remote Studio Data

One of the goals of implementing the phase 1 demonstrator is to create an environment at the studio premises to gather video and skeleton data from Kinect of the sign interpreter who is signing to the daily news broadcast contents. The media gateway component located at the remote studio is connected to the Blackmagic studio camera and to the Kinect v2 sensor to collect the video feeds, depth feeds and skeleton information which are to be transmitted to the processing and rendering unit (see D2.1 for detailed information). In addition, the media gateway is also assigned to with the task of storing these data in the local hard drives on a daily basis. The data collected in such manner will be then used in WP4 for deterministic grammar modelling and automatic sign language translation in phase 2. Therefore, a use case for collection of data at the remote studio in the media gateway is crucial for the upcoming tasks in WP4, and WP5 and is defined as below.

Use case 1.1: The signing data from the remote studio need to be captured and stored as training data for sign language recognition and translation.

Preconditions:

- Blackmagic 4K studio camera is connected to the media gateway.
- Kinect v2 sensor is connected to the media gateway.
- Precision clock is placed in the remote studio to be visible to both Kinect and studio camera.
- Specify the compression ratio, encoder settings for the required quality level.
- H.264 encoder instance is running expecting media data from Blackmagic studio camera, and Kinect sensor as inputs.

Postconditions:

- 2160p/50fps studio camera streams and Kinect data (RGB colour/25fps, depth/25fps, skeleton data/25fps) are synced, compressed and stored in the remote studio.

Main success scenario:

1. Synchronized studio camera feed, Kinect streams and skeleton data are stored in the remote studio's storage devices on a daily basis.

3.2. Transmission of Media Data from Remote Studio to Processing and Rendering Unit

The studio camera feed and Kinect streams from the remote studio that captures the remote sign interpreter are collected by the media gateway located in the remote studio. These media streams should be synchronised and transmitted to the processing and rendering unit for sign handshape recognition and 3D model rendering functions (see D2.1). The functionalities of processing and rendering unit depends on media gateway providing a continuous stream of media data and Kinect skeleton data at the specified format and frame rate defined in D2.1. In this case, it is imperative that use cases are defined for format conversion which are defined as follows.

Use case 2.1: The high frame rate video stream from the studio camera (2160p/50[1]) is down sampled and converted to 1080p/25[2].

Preconditions:

- 2160p/50 video feed is received from the Blackmagic SDI to the down sampling and rate conversion sub-component.
- Input port in ffmpeg encoding and muxing instance is open and listening for the video feed from the studio camera.

Postconditions:

- 2160p/50fps video is converted to a 1080p/25 stream and forwarded to the ffmpeg input port.

Main success scenario:

1. Media Gateway receives the downsampled video stream and is able to process and transmit it to the Media Receiver.

The media streams collected by the media gateway are initiated from two different sources (i.e., studio camera and Kinect). The deterministic grammar models and sign language translation algorithms and corresponding deep learning based training frameworks require these streams to be synchronised to a common timestamp. Therefore, a use case associated with the media gateway is defined as follows to handle the media synchronisation.

Use case 2.2: The studio camera stream, and Kinect streams (RGB colour, depth, and skeleton) are synchronized to the same timestamp.

Preconditions:

- The video feeds from all sources is available at a fixed frame rate.
- The input buffers for the media synchronizer is ready and available.

Postconditions:

- A synchronized stream of videos streams from multiple sources is available for compression and transmission.

Main success scenario:

1. The various streams are synchronized and available for processing and transmission to the media receiver.

Once media streams and Kinect skeleton data are available in the media gateway, they need to be prepared in order to be transmitted to the processing and rendering unit. Therefore, use cases are defined for content preparation, encoding, muxing and streaming as given below.

[1] The 2160p studio camera will capture high resolution images at a high framerate (50 fps) in order to ensure high quality training data is available for the subsequent phases (2,3) of the project.

[2] Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcast and Broadband Applications, DVB Document A001, July, 2017.

Use case 2.3: The video streams from the studio camera, Kinect RGB and depth streams are encoded and muxed into mpeg-ts container with timestamps for streaming.

Preconditions:

- Input ports of ffmpeg encoding and muxing instance are open and keep on listening to the video streams from the studio camera, and Kinect sensor.
- Output port of ffmpeg encoding instance is successfully connected to the media receiver in the processing and rendering unit.
- 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.

Postconditions:

- The H.264/AVC compressed video streams are available for streaming from media Gateway to media receiver in the processing and rendering unit.

Main success scenario:

1. Media receiver in the Processing and Rendering Unit receives a continuous stream of video data from the remote studio.

Use case 2.4: Media streams are streamed from the media gateway to the media receiver at the processing and rendering unit.

Preconditions:

- TCP server port in the media receiver is open and listening to receive media data.
- Connection is established between the media gateway and media receiver.
- A continuous stream of encoded mpeg-ts media data in available in the encoder buffer ready to be streamed.

Postconditions:

- A continuous stream of media data is available in the media receiver at 25fps.

Main success scenarios:

1. Handshape analyser receives a continuous stream of media data.
2. Handshape analyser outputs a handshape index to 3D model renderer.
3. 3D model renderer receives a continuous stream of media data.
4. 3D model renderer outputs an animated model.

Use case 2.5: The Kinect skeleton data are prepared as a frame based JSON stream to be streamed to the media receiver in the processing and rendering unit.

Preconditions:

- Kinect skeleton data is available in the frame buffer at 25fps.
- JSON parser is running accepting the data stream.
- Timestamps from the video frame encoding instance is available for the respective frames.

Postconditions:

- A JSON stream of skeleton data with frame numbers, and timestamps is available for streaming.

Main success scenario:

1. The skeleton JSON stream is synchronized with video streams and ready for transmission.

Use case 2.6: Kinect skeleton data are streamed from the media gateway to the media receiver in the processing and rendering unit.

Preconditions:

- TCP server port in the media receiver is open and listening to the Kinect skeleton data.
- A JSON stream of skeleton data is available in the output buffer of the media gateway.

Postconditions:

- A continuous stream of skeleton data is available in the media receiver at 25fps.

Main success scenarios:

1. Media receiver receives the Kinect skeleton data synced with the video feeds and forward it to the Handshape analyser
2. Handshape analyser receives a continuous stream of Kinect skeleton data
3. Handshape analyser outputs a handshape with high confidence

3.3. Transmission of Media Data within the Processing and Rendering Unit

Once media data and Kinect skeleton data 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 3.1: Studio camera feed, Kinect colour, depth and skeleton data are transmitted to the handshape analyser.

Preconditions:

- Media receiver in the processing and rendering unit receives the video feeds and Kinect skeleton data.
- TCP server ports are available in the handshape analyser to receive the video and skeleton data.
- Media server is successfully connected to the handshape analyser.

Postconditions:

- Handshape analyser receives a continuous stream of images and skeleton data.

Main success scenarios:

1. Handshape analyser is running continuously to predict the handshape.
2. Handshape analyser outputs a handshape index.

Use case 3.2: Studio camera feed, Kinect colour, depth and skeleton data are transmitted to the 3D model renderer.

Preconditions:

- Media receiver in the processing and rendering unit receives the video feeds and Kinect skeleton data from the media gateway.
- TCP server ports are available in the 3D model renderer to receive the video and skeleton data.
- Media receiver is successfully connected to the 3D model renderer.

Postconditions:

- 3D model renderer receives a continuous stream of images, depth and skeleton data.

Main success scenario:

1. Handshape analyser outputs a handshape index to the 3D model renderer.
2. 3D model renderer continuously renders and outputs the rendered images at 25fps.
3. C4A 3D model displays the same facial expression as the remote sign language translator.

The 3D model renderer suffers from the limitation of recognizing and rendering the handshapes directly from the media streams received from the media gateway in the remote studio. Hence, handshape analyser component first analyses the video data and Kinect skeleton data and determines the handshapes (see section 3.4) 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 handshape index details to the 3D model renderer.

Use case 3.3: The estimated handshapes are transmitted to the 3D model renderer.

Preconditions:

- Handshape analyser determines the handshape.
- A JSON stream of handshapes with frame numbers are available.
- Handshape analyser is connected to the 3D model renderer through a TCP socket to transmit the data.

Postconditions:

- Handshape for each frame are available at the 3D model renderer.

Main success scenario:

1. Handshape analyser outputs a handshape index to the 3D model renderer.
2. 3D model renderer receives and renders the handshape in real-time.
3. 3D model renderer continuously renders and outputs the rendered 3D model images at 25fps.

Once the 3D model renderer receives the media data and Kinect data from the remote studio and handshape details from the handshape analyser, it renders the 3D model corresponding the remote signer (see section 2.4.5). Once 3D model images are rendered, they need to be mixed with the original broadcast stream 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 3.4: The rendered 3D model image is transmitted to the mixer/encoder for overlaying and mixing with the main broadcast stream.

Preconditions:

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

Postconditions:

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

Main success scenario:

1. Encoder/mixer continuously mixes the rendered images with the broadcast stream at 25fps.
2. Encoder/mixer continuously encodes the mixed video in multiple profiles and output them as MPEG-TS.

3.4. Estimating the Handshape based on the Remote Studio Video/Skeleton Feeds in the Processing and Rendering Unit

The first demonstrator in M12 focuses on rendering a 3D model of the remote sign interpreter into the broadcast stream in order to facilitate the sign language representation to the broadcast content. The handshapes, body pose and facial expressions of the human sign interpreter need be rendered onto the 3D model to achieve a realistic representation of the sign interpreter and the sign language representation of the content. Therefore, handshape analyser component is proposed to support the 3D model renderer to determine the handshapes which need to be rendered. In this context, the following use cases are defined to describe the functionalities of the handshape analyser.

Use case 4.1: Handshape analyser identifies the location of the hands using Kinect skeleton stream.

Preconditions:

- Handshape analyser receives an RGB image of the studio camera feed.
- Handshape analyser receives the Kinect streams including the skeleton data.

Postconditions:

- The locations of the hands in the video feed are identified.

Main success scenario:

1. Handshape analyser accurately determines the corresponding handshape index.
2. 3D model renderer accurately renders the correct handshape on to the model.

Use case 4.2: Extract texture of the hands using image patch extractor sub-component.

Preconditions:

- Handshape analyser receives the RGB images from the remote studio via media gateway and media receiver.
- The location of the hands is accurately identified.

Postconditions:

- An image patch containing the handshapes is extracted from the HD video streams.

Main success scenario:

1. Handshape index is estimated from the Deep Neural Network (DNN) inference module.
2. 3D model renderer gets the handshape indexes at 25fps.

Use case 4.3: Frame buffer is filled with handshape image patches as a stream for a duration of 10s.

Preconditions:

- Image patch extractor, extracts the handshapes from the HD video stream images.
- Buffer has sufficient space for the incoming handshape images.
- Buffer maintains a server port listening to the handshape images.

Postconditions:

- DNN receives a continuous stream of handshapes for analysis.

Main success scenario:

1. Handshape analyser makes handshape estimations with a higher accuracy compared to the single frame based prediction scenario.
2. 3D model renderer renders the accurate handshape with a higher confidence level.

Use case 4.4: Deep Neural Network inference module classifies the handshape with a confidence level.

Preconditions:

- DNN is loaded with the accurate prediction models.
- DNN receives a stream of handshapes or individual handshapes from the stream buffer.

Postconditions:

- DNN makes a prediction on the handshape.

Main success scenario:

1. Handshape index is transmitted to the 3D model renderer along with the frame number.
2. 3D model renderer renders the handshape on the 3D model.

Use case 4.5: Handshape index is transmitted in JSON format to the 3D model renderer with the frame index.

Preconditions:

- DNN makes a classification on the handshape.

Postconditions:

- JSON string is prepared with the handshape ID and frame number.

Main success scenario:

1. 3D model renderer renders the handshapes continuously in real-time without any delays.
2. 3D model is rendered on to the broadcast streams without any delays/jitter.

3.5. Rendering of 3D Model of the Sign Interpreter

For pilot 1 a remote-studio will be available where a TV-camera and an RGB-D camera will capture a human sign language translator. The captured data will be transmitted to rendering module, where control commands for the C4A 3D model will be extracted from the remote data. While the RGB-D sensor will deliver skeleton joint positions, we have no exact information about the head pose and orientation. Since the head pose is a key information for a successful facial expression transfer (from the sign language translator to the 3D model) it is necessary to extract the head pose information from the received image and depth data. There, a use case is defined.

Use case 5.1: Real-time estimation of the sign language translator's head pose based on the remote studio data.

Preconditions:

- 3D model renderer receives the RGB and depth images as well as skeleton data from the remote studio via media gateway and media receiver.

Postconditions:

- The rigid pose of the sign language translator's head was computed.

Main success scenario:

1. The C4A 3D model is rendered with a realistic facial expression.
2. The C4A3D model is rendered with the sign language model's facial expression.

Another important part of the rendering module is the animation of hand and fingers. The 3D model renderer will receive a stream of recognized hand shapes from the handshape analyser and display it. Since the detected handshapes will refer only to a discrete set of approximate 60 canonical shapes, it is necessary to create a smooth animation between the detected hand shapes. Therefore, a realistic animation path must be created that allows displaying natural hand and finger movements which is defined in the use case below.

Use case 5.2: Generation of an animation path for hand/finger animation.

Preconditions:

- 3D model renderer received a new target handshape with timing information from the 3D model language translator.

Postconditions:

- A sequence of joint angles was generated to display the target hand shape.

Main success scenario:

1. 3D model renderer renders the C4A 3D model with the desired handshape at the given time point.

2. The C4A 3D model shows a natural hand/finger movement to reach the desired handshape.

Facial expressions are critical for the transfer of information in sign language therefore the 3D model must be able to display the facial expression of the remote sign language interpreter. This will be performed by extracting a dynamic texture from the video stream using the estimated head pose of the sign language interpreter. Knowing the head pose it is possible to project the video frame into texture space and update the 3D model's facial texture from it. The following use cases is defined.

Use case 5.3: Facial expression transfer for facial animation.

Preconditions:

- 3D model renderer computed the head pose of the sign language interpreter.
- 3D model renderer receives the RGB and depth images from the remote studio via media gateway and media receiver.

Postconditions:

- The C4A 3D model's facial texture is update using the live image by projecting it into texture space.
- The new facial texture is integrated seamlessly into the 3D models texture.

Main success scenario:

1. The C4A 3D model shows the facial expression of the remote sign language interpreter.

The skeleton information, which is provided by the RGB-D sensor, cannot be used directly to animate the 3D model. Therefore, the skeleton tracking information must be transferred onto the 3D model's skeleton such that the 3D model is performing the desired body movements. Therefore, a use case is defined.

Use case 5.4: Body animation from skeleton information.

Preconditions:

- The 3D model renderer receives the skeleton data from the remote studio via media gateway and media receiver.

Postconditions:

- The C4A 3D model's new body pose reflects the body pose provided by the received skeleton data.

Main success scenario:

1. The C4A 3D model shows the body pose/movements of the sign language interpreter.

As the system is intended to be used for "live" translation in a TV studio setting it is necessary to achieve real time in animation and rendering, for what a use case is defined.

Use case 5.5: Real-time animation of the C4A 3D model based on the remote studio data.

Preconditions:

- 3D model renderer receives the RGB images as well as skeleton data from the remote studio via media gateway and media receiver.
- Sign language translator's head pose was estimated from the remote studio data.
- 3D model renderer receives the handshape data from the handshape analyser.
- 3D model renderer updates the 3D model texture based on the estimated head pose.

Postconditions:

- 3D model geometry changed to represent the current pose and hand shape of the remotely captured sign language translator.
- 3D model texture is updated to represent the remote sign language translator's facial expression.

Main success scenario:

1. 3D model renderer creates a realistic image of the C4A 3D model showing the same pose and facial expression as the remote sign language translator.
2. The system is able to process the incoming data in real time and generates a video output approximately 25 fps.

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

It is difficult, and in some cases impossible, to mix two separated streams on a client application, due mainly to client device limitations (missing or weak HW capabilities). Thus, in order to enable the client applications rendering the 3D model beside the main video content, the system will mix two separated video streams server-side, in the processing and rendering unit. Before mixing the two video streams, they should be synchronized, even if it is not required a strict synchronization down to the single frame; the system will apply a fixed delay to the main video content, in order to compensate the delay generated by the remote studio transmission (only in phase 1) and by the 3D model generation process (in phase 2).

After mixing, the system will encode the resulting video stream, 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 6.1: The video stream from the broadcaster is synchronized to the 3D model stream.

Preconditions:

- The video feeds from broadcaster is available.
- The input buffers for the media synchronizer is ready and available.

Postconditions:

- A synchronized stream of video is available for decoding and mixing.

Main success scenario:

1. The video stream from the broadcaster and the 3D model are synchronized and ready to be mixed together.

Use case 6.2: The video streams from the broadcaster and the rendered 3D model stream are mixed and encoded into mpeg-ts container.

Preconditions:

- Input ports of ffmpeg encoding and mixing instance are open and keep on listening to the video streams from the broadcaster stream and the 3D model stream.
- 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.

Postconditions:

- 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 where the 3D model stream is mixed along with the Broadcaster video stream.
2. The origin server delivers the mixed video stream to the user terminals.

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

In order to transmit the mixed 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 7.1: The HbbTV app and user's devices web app download the mixed video stream from the origin server.

Preconditions:

- 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.
- Web application on user's devices is ready to request video stream chunks.

Postconditions:

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

Main success scenario:

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

3.8. Business Models

Different business models for the C4A system phase 1 demonstrator are defined that will remove barriers for media consumption. They address political and governmental issues arising from the UN Convention on the Rights of Persons with Disabilities, ratified in 2014, especially articles 30 and 21.

Therefore, besides TV production other application areas like

- Government and Political Parties (parliamental and political debates live on TV or Internet),
- Events (large scale events, often available on TV) and
- Education (teaching at school, university and educational institutes in general),

potentially generate demand for C4A technology. As this deliverable concentrates on the phase 1 demonstrator, the use cases only address broadcaster's TV production as applicable business area. All business models are described in detail in "D6.3 Business Model Analysis", chapter 3. Hence, to avoid repetition, D2.2 just lists the use cases for business models.

Use case 8.1: The C4A system meets the demands of the UN Convention on the Rights of Persons with Disabilities, especially article 30 and article 21.

Use case 8.2: The C4A system supports broadcasters sign language TV content via broadcast side mix (HbbTV 1.4 and above) and user side mix (HbbTV 2.x) to provide the sign language service to as many people as possible but also show the potentials of the advanced user experience based on the HbbTV 2.x technology.

3.9. Use Cases and Matching C4A Objectives

The use cases that are identified in previous sections fulfill the following C4A main objectives, listed in the following table.

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

Project Objectives	Matching use cases
1. Sign-interpretation parameter extraction, 3D model generation and real-time animation	5.2, 5.3, 5.4, 5.5
2. Signal processing for sign-language interpretation	1.1, 4.1, 4.2, 4.3, 4.4, 4.5, 5.1, 7.1
3. Experimental performance evaluation and creating sustainable business models	8.1, 8.2
4. Building a large-scale and marketable demonstrator	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 3.1, 3.2, 3.3, 3.4, 6.1, 6.2, 7.1, 8.1, 8.2

3.10. Risk Analysis

Several risks and challenges regarding the realization of C4A use cases due to dependencies among components, tasks, interactions and inter-relationships do exist. The following table lists risks that may arise during C4A system development associated mitigation strategies.

Table 2: Initial C4A Risk Analysis

Risk	Likelihood	Impacting use cases	Mitigation plan
Recognition accuracy is poor in the handshape analyser	High	4.1, 4.2, 4.3, 4.4, 4.5	Additional visual cues can be used to help resolve ambiguity. Temporal smoothing can be employed. Number of classes can be reduced.
Real-time operation of decoders and demuxers in media receiver becomes an issue	Low	2.1, 2.3	Increase the number of processing cores allocated to the operations. Utilise GPU assisted hardware accelerated decoding
Packet losses due to buffer over flow or network overflow	Low	2.3, 2.4, 2.5, 2.6, 4.3, 4.4, 6.1, 6.2	Use TCP sockets for data exchange, which guarantees no packet losses. Increase the decoder buffer sizes to accommodate raw video frames in case of low frame rate processing of 3D model renderer and handshape analyser
TCP ports become unavailable for communication	Low	2.4, 2.6, 3.1, 3.2, 3.3	Perform TCP port availability checks before the program initialisation.

Risk	Likelihood	Impacting use cases	Mitigation plan
Real-time software encoding of high quality video streams may not be possible in the Media Gateway	Low	1.1, 2.3, 2.4, 2.6, 4.2, 5.1, 5.5	GPU assisted hardware accelerated encoding will be considered. Video quality will be adjusted to maintain the expected encoding rate.
Errors in timecode base video synchronisation between Kinect and Blackmagic studio camera	Medium	1.1, 2.2	A time-clock will be made visible to both inputs, and a manual synchronisation is achieved by extracting the visible timecodes.
8-bit depth data is insufficient for accurate handshape recognition	Medium	4.1, 4.2, 4.3, 4.4, 4.5, 5.2, 5.5	16-bit depth data will be streamed in two different 8-bit channels and combined at the media receiver.
Real-time operation of decoders and encoders in mixer and encoder component becomes an issue	Low	6.2	Increase the number of processing cores allocated to the operations. Utilise GPU assisted hardware accelerated decoding/encoding.
Storage available space containing the registered video stream and the skeleton Kinect data is finished	Low	1.1	Increase the size of the storage space
Network bandwidth isn't enough to stream content between remote studio and processing and rendering unit	Low	2.1, 2.2, 2.3, 2.4, 2.5, 2.6	Increase the bandwidth of the network interfaces. If not possible, decrease the quality of the video stream.
Low facial tracking accuracy.	Medium	5.1, 5.3, 8.1	Run tracking routine on GPU for faster & more tracking iterations Add light compensation to improve the optical flow based face tracking Use automatically detected facial features for consistency checks and re-initialization
3D model shows unnatural body poses	Medium	5.4, 8.1	Implement additional constraints to filter wrong skeleton configurations that are provided by the RGB-D sensor
The model renderer does not achieve real time performance	Low	5.1, 5.3, 5.5, 8.1	Run time critical algorithms on GPU in order to improve the speed
Remote sign interpreters facial texture does not match with the 3D model's appearance/skin.	Medium	5.3, 8.1	Perform colour compensation to match the extracted facial texture with the 3D model's facial texture.

4. Requirements

In this section, requirements are specified, which define the characteristics of the C4A system, to be satisfied by the C4A system within phase 1 demonstrator. The requirements face the wide scope of user requirements, which can, also like use cases, be assigned to the four main C4A components. Section 4.1 addresses the remote studio- and broadcaster-requirements, sections 4.2 until 4.4 the rendering and processing unit and its interaction to the remote studio. Sections 4.5 faces requirements the users have regarding the users' terminal, section 4.7 deal with business models. The chapter ends with the matching of the requirements to the C4A objectives.

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: Initial C4A Broadcaster Requirements

Requirement ID	Req. description	Methodologies followed to fulfil the requirement
REQ1.1	The video content stream must be sent to the Remote Studio.	To stream the video content through the Remote Studio we leverage on existing standard Broadcaster systems like DVB Playout.
REQ1.2	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.
REQ1.3	A storage device must be present to collect video and skeleton data for analysis.	Video stream of Studio camera and Kinect skeleton data will be stored on storage devices within the broadcaster premises.

4.1.2. Remote Studio Requirements

Table 4: Initial C4A Remote Studio Requirements

Requirement ID	Req. description	Methodologies followed to fulfil the requirement
REQ1.4	High quality images of the remote sign interpreter are required to extract the hand shape and facial expressions for sign language recognition and translation.	The 4K (3840x2160) images captured by the Blackmagic studio camera at 50fps, will be compressed using H.264/AVC with crf 20 encoder

Requirement ID	Req. description	Methodologies followed to fulfil the requirement
		setting, and stored in the storage facilities in remote studio.
REQ1.5	Extraction of the accurate hand position, body pose is required for the handshape recognition.	The Kinect sensor is used to extract the skeleton data, depth details at 25fps along with the Kinect's RGB images (25fps) and are stored in the storage facilities in remote studio.
REQ1.6	The data streams from the Kinect and 4K video streams from the studio camera needs to be synchronized.	A precision clock will be positioned in the remote studio which is within the field of view of both Kinect and studio camera. Both studio camera feed and Kinect feeds are collected at the "Media Gateway" and are synchronized to the same timestamp before compressing and storing them.

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 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: Initial C4A Model Creation and Capture Requirements

Requirement 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. Since the C4A 3D model hands/fingers are controlled on a key-frame basis, it is necessary that the movements blend seamlessly into each other. Finally, in order to achieve an automatic translation that can be used in live programs, it is necessary that the animation as well as the rendering can be performed in real time although a delay would be acceptable. The following requirements are derived.

Table 6: Initial 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 C4A 3D model should be able to interpolate hand gestures seamlessly between different (key) -poses.
REQ2.14	The animation of the C4A 3D model should be achieved in real-time.
REQ2.15	The rendering of the C4A 3D model should be achieved in real-time.
REQ2.16	Handshape analyser receives a continuous stream of synchronized video data, Kinect skeleton data from the remote studio.
REQ2.17	Motion blur needs to be avoided in the video streams this will be the largest source of errors in the hand shape analysis.
REQ2.18	Handshape analyser predicts the sign handshape and provides a handshape index with the frame number to 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.

This section specifies the networking requirements for phase 1 demonstrator which covers the end-to-end content network monitoring and methodologies to fulfil them.

Table 7: Initial C4A Networking Requirements

Requirement ID	Description	Methodologies followed to fulfil the requirement
REQ3.1	The synchronized video feeds from the studio camera, Kinect RGB, Kinect depth and skeleton data, needs to made available in the processing and rendering unit for handshape analysis and 3D model rendering.	Media gateway located in the remote studio is configured to collect the video streams from multiple sources, synchronize them, encode and packetize the data and stream to the media receiver located in the processing and rendering unit.
REQ3.2	The video streams transmitted to the processing and rendering unit for handshape analysis and 3D model rendering should be sufficient quality.	The level of quality required is determined in WP3 and WP4. An approximate bandwidth of 100Mbps is maintained between the media gateway and media receiver in the processing and rendering unit.
REQ3.3	Two HD streams, and 1 Kinect depth stream needs to be compressed in real time for 25fps and one 4K stream need to be compressed in 50fps.	Hardware accelerated software encoding is performed using NVIDIA GPUs, and ffmpeg encoding libraries. Extra CPU cores are allocated to the media gateway to avoid processor overload.
REQ3.4	Handshape analyser should receive the synchronized Kinect RGB images, studio camera images in RGB, and Kinect skeleton data at 25fps.	Media receiver receives the compressed streams of Kinect colour images, studio camera images, and skeleton data. Media receiver decodes them, and forwards to handshape analyser through TCP sockets.
REQ3.5	3D model renderer should receive the synchronized Kinect RGB images, Kinect depth images, studio camera images in RGB, and Kinect skeleton data at 25fps.	Media receiver receives the compressed streams of Kinect colour/depth images, studio camera images, and skeleton data. Media receiver decodes them, and forwards to 3D model renderer through TCP sockets.
REQ3.6	3D model renderer should receive the handshape, facial expression, mouthing information for a particular signing sequence.	Handshape analyser outputs the frame ID, timestamp, handshape index, as a JSON string, which is transmitted to the 3D model renderer as through a TCP socket
REQ3.7	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 a IP stream through TCP sockets

Requirement ID	Description	Methodologies followed to fulfil the requirement
REQ3.8	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
REQ3.9	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
REQ3.10	The mixed frames need to be made available at the encoder to real-time encoding and preparation of 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.
REQ3.11	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: Initial C4A Encoding, Mixing and Streaming Requirements

Requirement 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 includes a basic understanding of the requirements of the deaf community with respect to the domain of TV. To understand the background for the requirements of the deaf user group regarding the C4A system, this chapter provides introducing information about the deaf community (section 4.5.1), e.g. the communication patterns, sign language and its grammar structure. Furthermore, this chapter gives an overview about best practise examples of TV shows with sign language, especially for the key points TV interface design (section 4.5.2) and the signer (section 4.5.3). Out of these information, initial requirements will be deduced at the appropriate passages.

All information of this chapter is based upon first interviews with deaf signer and reviewing various literature. This deliverable faces initial requirements, the interviews are still running. One interviewee is an expert in the field of TV interpretation for the Deaf. However, these views do not reflect the opinion of the entire deaf community. Nevertheless, some literature review has been utilized to back these suggestions.

4.5.1. Deaf People and Sign Language

Attention to the usability and user experience of C4A needs to be given right from the early stages. For the same, it is necessary to collect data on the requirements of the user which should be employed during the development process. To understand such user requirements at an early stage, the first most important aspect is to understand how the users communicate (as television is a form of one-sided communication) among each other and to hearing people. Deaf people understand their surroundings only through visual input. Their mother tongue is sign language. Written language such as subtitles for instance is equivalent to a foreign language for them.

Most deaf people prefer to communicate in their mother tongue sign language. This enables them to effectively communicate with a limited number of people such as other deaf, friends and family who can sign and people who use signing for their profession such as interpreters for instance. Another possibility is to communicate in written form if necessary. Deaf people make use of various digital communication methods such as teletypewriters, emails, SMS, chats such as ‘WhatsApp’ etc. Email is the most widely preferred, with SMS and other chat forms are today more common among the younger deaf population

(Wilson & Hoong Sin, 2015). Video chats like 'skype' and 'facetime' have made it possible to communicate over a distance in sign language. When communicating with hearing counterparts who can't sign, deaf people often make use of interpreters. Also, a small percentage use lip reading and their own voice to communicate.

Kurz & Mikulasek (2004) state that those who are born deaf prefer to watch television in their mother tongue i.e. sign language whereas those with acquired hearing loss prefer to watch television with captions as there are not very familiar with sign language. Broadcasters prefer subtitling over the use of sign language. The major reason is that captioning is more cost effective and that subtitling can reach a larger majority as compared to signing which is accessible to only a small group i.e. the deaf community. The deaf community however does not agree to this view, they are of the opinion that signing should be recognized (Kurz & Mikulasek, 2004). Theye prefer to view TV in their first language. Subtitles are written languages and have a different grammar structure as compared to sign language, hence, making it difficult for them to follow.

Sign language is a visual language. It is clear that as any language, sign language has evolved over time. Differences can be found between signing in the same country let alone around the world. These regional variations are however not as large as the international differences (Kyle & Woll, 1988). There are 148 sign languages (e.g. American Sign Language (ASL), British Sign Language (BSL)) plus many dialects. This makes it difficult and inappropriate to display content for any content with a global audience. For this reason, it is necessary to make localized variants for all media content so that such media content is accessible to all deaf people (Brewer et al., 2015). Therefore, we deduce initial requirement REQ5.1 for C4A system, that are listed in Table 9.

"Sign language is a complex combination of facial expressions, mouth/lip shapes, hand and body movements, and finger spelling" (Muir & Richardson, 2005, p.1). Signs are produced by gestures of the hand, arms, face and body (Liddell, 2000). Gestures, hand movements and positions, finger movements and positions, facial expressions, (that include mouth, cheeks, eyes, eyebrows) body movements, position and attitude etc. are all part of sign language. Muir & Richardson (2005) found that the deaf people look at the face of the signer concluding that the facial expressions of the signer is very important in giving clues to the meaning of the gestures and hand/finger movements used during signing. Some signs/gestures may have a similar meaning in different countries/languages but mouth movements are different. Dialects within the country are often similar. Mouth movements in a particular sign language are fixed. Eyes and body position normally don' t change according to countries.

This form of visual communication can often be very rapid. Which means freely expressed sign language communication is a detailed combination of rapid hand (especially finger) and body movements and facial expression. Such rapid movements, for example during finger spelling, can be seen as a blur and must be hence follow some quality requirements when recorded or captured (Muir & Richardson, 2005). As described above, sign language needs several of components. Therefore, we propose initial requirements REQ5.2 – REQ5.5 for C4A system (see Table 9).

Sign language grammar is different from spoken language grammar. A basic sentence in written English language constitutes of a subject, a verb and then the object. In sign language you first mention the subject, then the object and then the verb. Verbs and pronouns are shown by directing the sign towards the subject/object.

Sign language is often abbreviated which means it makes use of a fixed sign or set of signs to convey specific messages and not word to word signs which is often very long and time consuming. Sign language does not make use of articles. Idioms also have a particular a sign or set of signs. Earlier, people made use of word to word translation. This changed from the nineteen hundreds when the grammar rules were extensively used. Older deaf people however still make use of word to word sign translation. Exemplary sentences in written language and in sign language are listed below.

Examples	English	Sign Language
	I drive a car.	I car driving.
	I give you the/some bread.	I bread give. (Directed towards the person which indicates pronoun)

I am not interested to clean the house.	I house clean no interest.
The mouse runs away quickly.	Mouse quickly run away. (The sign for this sentence is shorter than the signs for each word. When each word is signed individually, it is long and takes a lot of time.)

Therefore, we deduce initial requirements REQ5.6 and REQ5.7 (see Table 9).

Signs can make use of one hand or two hands. Certain words need to be signed with two hands while others with one.

Positions and Locations of the hands are important, too. Some signs make contact with a facial or body part while some are made in the air/space near a facial or body part. Some signs have the same finger position or gesture. The sign for “point” and “goal” are the same. The position of the sign distinguishes the two words. E.g. sign for “goal” is on the level of the forehead, whereas the sign “point” is on the level of the abdomen. If the position is changed, it could result in another sign or no sign (Liddell, 2000).

Certain words such as new terms could not have a sign. When a sign for a word does not exist then each letter is signed. Tenses are showed with the help of signs and body position. For e.g. when conveying a message in future tense, first “future” is signed followed by the content and in past tense first “past” is signed followed by the content. Signing people have a dominant hand. It can depend on if the person is left-handed or right-handed. Right- handed people primarily sign with their right hand, left-handed with left. This can course be interchanged -depending on the situation.

The location of the subject/object also matters with which hand is being used. Location also influences the position of the upper-body which needs to be turned towards the object/subject.

Based on the key points mentioned above, initial requirements REQ5.8 – REQ5.12, for C4A TV interface design are compiled (see Table 9).

Table 9: Initial C4A User Requirements – General System

Requirement ID	Req. description
REQ5.1	The system should display TV content to deaf watchers in their national sign language.
REQ5.2	C4A 3D model should comprehend a perfect and timely combination of signs and gestures with facial expressions.
REQ5.3	C4A 3D model should include signs and gestures in coordination with its respective mouth position and movements to avoid misinterpretation.
REQ5.4	C4A 3D model should display signing at a normal speed i.e. signing should be freely expressed and not slowed down. Signing and gesturing should not be exaggerated.
REQ5.5	Sufficient video quality of the sign translated information is required.
REQ5.6	C4A system must display sign language in their correct grammar.
REQ5.7	Body position should be clearly portrayed.
REQ5.8	Body position of the 3D model should be clearly portrayed.
REQ5.9	C4A 3D model should possess precise hand and finger coordination.
REQ5.10	C4A 3D model should include exact hand and finger positioning.

REQ5.11	C4A 3D model should include exact hand locations.
REQ5.12	Overall quality of signing as well as sign translated content should be clear and precise without any mistakes.

4.5.2. Sign Language in TV Shows

It is important to consider the on-screen presentation of the content which will normally include two visual inputs, the sign language interpreter and of course the actual content of the program. Sign interpreters are commonly presented in a box (or in an oval or egg-shaped, a cloud etc.) using picture-in-picture technology or using chroma key technology (see Figure 3 and Figure 4).

The use of picture-in-picture was earlier more common however studies report that this method is not optimal due to insufficient details. Viewers of sign language TV prefer the sign interpreter to be larger as displayed in figure 3, thus increasing clarity (Orero et al., 2014). The main content at Figure 5 does not occupy the entire screen but contains a shared background with the sign language interpreter. The interpreter further overlaps this content (Orero et al., 2014).

Background with a contrasting colour to the cloths of the interpreter is preferred as the hand movements are better visible. A deaf sign interpreter who is an expert in his field suggests the use of a plain background that is in contrast to the colour of the content. Textured backgrounds are to be avoided. He also suggested that content is best presented in a trapezoid and not a rectangle as in Figure 5. Therefore, we deduce initial requirements REQ5.13 – REQ5.15, regarding the C4A TV interface design and listed in Table 10.

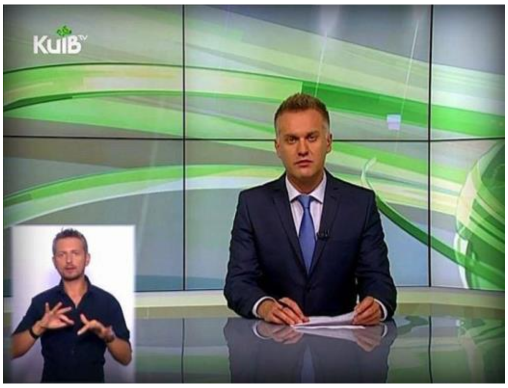

Figure 3: Picture-in-picture technology (source: https://upload.wikimedia.org/wikipedia/commons/c/cb/TV_Interpreter.jpg [23.01.2018])

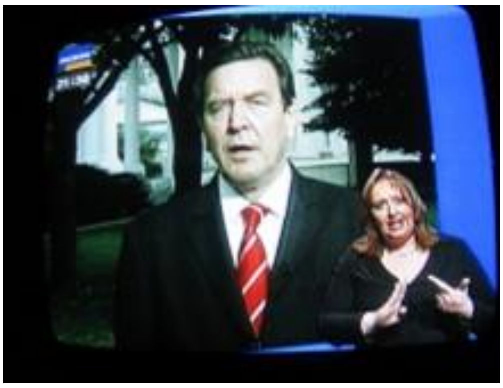

Figure 4: Chroma key technology (source: https://www.flickr.com/photos/bsktcase/49152259 [23.01.2018])

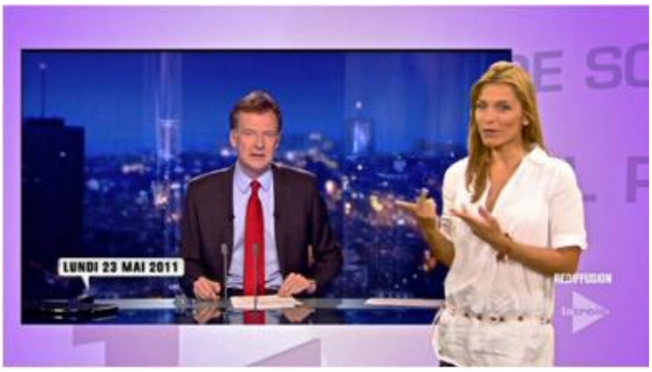

Figure 5: Signing on RTBF, Belgium (Orero et al. (2014), p. 190)

Table 10: Initial C4A User Requirements - GUI Design

Requirement ID	Req. description
REQ5.13	Background colour should be contrasting the cloths of the 3D model as well as the content.
REQ5.14	Background should be plain. Textured backgrounds should be avoided.
REQ5.15	Content should be presented in a trapezoid.

4.5.3. Sign Language Interpreter in TV Shows

> *"It is believed that with further research animated 3D model may be a sufficient and more practical option in the future"* (Brewer et al., 2015).

For a signing 3D model in project phase 2, it is preferred that it is portrayed in the form of a realistic human adult. A realistic child 3D model is acceptable only for children shows. This animated interpreter should wear dark cloths. Strips are forbidden as it is often difficult to see the hand movements. Long sleeves are recommended. The 3D model must not have a beard as it is challenging to recognize mouth movements.

Wochenschau, a 25 minutes program was aired in the 1990's in the late hours. The research conducted before airing this program showed that the majority of the votes from the deaf community were in favour of a combination of sign language interpreting and captioning (Kurz & Mikulasek, 2004). In order to make Wochenschau realistic, precise coordination and close cooperation was required between the program makers and interpreters. The appearance and attire of the interpreter was deemed to be important. The paper mentions that plain-colored cloths like black, brown, grey etc. is preferred over colourful cloths. In addition, according to a survey, the deaf and hearing impaired people in Great Britain preferred signers who were pleasing to the eye, looked knowledgeable on the subject content that was signed, looked authoritative and who did not greatly contrast with the picture content such as its colour (Kurz & Mikulasek, 2004). Based on these key points, we deduce initial design requirements REQ5.16 – REQ5.22 (see Table 11) for the C4A 3D model in phase 1.

First interviews with the user group revealed that the most important feature is facial expressions and in particular the eyes of the 3D model (see also 4.5.1). Sign time made use of 3D model, but it lacked emotions and had no eyebrows resulting in imperfect facial expressions. Other important features of the 3D model should include eyebrows movements, cheeks movements, eye movements and mouth movements as well as mouth position. Deaf people sign in the area from the belly to a little above the head and from one elbow to another. A deaf interpreter recommended overall lighting to see facial expressions clearly.

In an experiment with eight deaf people, Muir & Richardson (2005) confirmed the findings that the face of the signer in perceived in high visual resolution and the hand movements are viewed in peripheral lower resolution vision. For deaf people, the face is the center of attention when viewing sign language, in particular when the signer uses less wide-ranging body gestures and more hand gesturing and finger spelling. It was also observed that when the video showed a close view of the signer, the participants gaze was mostly in the upper face region i.e. focusing more on and around the eyes of the signer. When the signer was further away on the screen and used wider gestures, the focus of the participant was on the lower facial region. It was also found that aspects like background distance of the signer from the camera and movement of the signer around the scene made no statistically significant difference to the viewing pattern (focus on facial region) of the deaf people. The gender of the 3D model matters according to content of the video. It also is good to change the 3D model according to the topic. Care should be taken to the fact that the timing of signing and that of video should match. Therefore, we deduce the initial requirements REQ5.23 and REQ5.24, regarding the C4A 3D model design.

Table 11: Initial C4A User Requirements - 3D Model

Requirement ID	Req. description
REQ5.16	C4A 3D model should look realistic and not as an animated character.
REQ5.17	C4A 3D model should be a human adult. Usage of child 3D models should be avoided with an exception to programs for children.

Requirement ID	Req. description
REQ5.18	C4A 3D model should wear colours in contrast to their skin colour for clear hand movement recognition. (When the 3D model has a lighter skin colour, the 3D model should wear darker shades. 3D model with darker skin colour should wear lighter shades).
REQ5.19	C4A 3D model should wear plain cloths. Striped, checked, colourful cloths should be avoided.
REQ5.20	C4A 3D model should wear long-sleeves.
REQ5.21	C4A 3D model should not have facial hair such as a beard.
REQ5.22	C4A 3D model should be pleasing to the eye, look knowledgeable on the content subject and look confident.
REQ5.23	C4A 3D model should include realistic facial expressions that is clear and identifiable as well as it displays perfect emotions. Therefore, the following components must be displayed realistic: eyes and eye movements, eyebrows and eyebrow movements, cheek movements, mouth movements.
REQ5.24	Facial movements should contain timely well-coordinated movements of eyes, eye-balls, eyebrows, cheeks and mouth.

4.6. Evaluation Tools

According to the usability engineering lifecycle, possible C4A users will participate in system evaluation. This section focuses on providing first information about possible 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 consist of 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 from 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). Vermeeren et al. (2010) stated that the relationship between usability and UX is intertwined. One important difference is that usability focuses on the task performance (execution time) whereas UX focuses on the lived experience (user's motivation and expectations). In comparison to usability, UX 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.

The UX of a system consists of two qualities: pragmatic or instrumental qualities and hedonic or non-instrumental qualities (Schrepp, Hinderks, & Thomaschewski, 2014). Instrumental qualities refer to product qualities such as its utility and usability. Non-instrumental qualities focus on aspects such as the product's aesthetics and its haptic quality i.e. the feel of the product. A third component refers to the emotional reaction of users that is stimulated by the product or system. Thus, the above-mentioned component results 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, when 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 user experience are closely linked with each other and to Usability. Quality of service refers to the technical characteristics that are related to the service 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 user experience helps in exploring how users feel about using a product, system or service (Vermeeren et al., 2010). As quality of experience and user experience focuses on the perception of the user, they can be said to share closer ties as both are related to measures received from users. Based on the key points mentioned above, the initial requirements REQ6.1 – REQ6.6 for C4A evaluation are compiled and listed in Table 12.

During the evaluation phases, which are repeated several times, 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-centered design process according to EN ISO 9241-210 and its conversion in several steps of requirement analysis and their iterative evaluation methods (utility, usability, user experience). Depending on the evaluation stage, different qualitative and/or quantitative (physiology) methods are used. The information acquired is then evaluated and processed. In the further course of project, it has to be examined which methods best fit for the evaluation of C4A system. Typical quantitative data what are imposed in laboratory settings are the efficiency, error rates and processing time in human's interaction with a system.

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.

A typical method for examining system's usability and UX, and what we focus for C4A evaluation phases, are also questionnaires. It collects the user's overall view-points, feelings, reactions on a particular product or service that was introduced to them. Well-known standardized questionnaires, such as UEQ (Laugwitz et al., 2008), AttrakDiff (Hassenzahl et al., 2003), SAM (Bradley & Lang, 1994) or SuS (Brooke, 1996), that assess overall UX can be utilized to gather the overall experience of the users for C4A. Such questionnaire also evaluates if the user experience of a particular service is sufficient i.e. if the UX of a service is sufficiently high so that it fulfils the general expectations of the users (Schrepp, Hinderks, & Thomaschewski, 2014). However, often it is better to take these questionnaires as a basis and to adapt them respective to the needs of the evaluation phase. Questionnaires can be qualitative and well as quantitative in nature. They can be applied flexibly in various evaluation environments, like in laboratory settings, field studies or other qualitative analyses. One advantage of questionnaires is that they can be administered to a larger population in a short period of time. This can be conducted traditionally via paper and pencil or even digitally/online. Common scales that assess UX are for e.g. affect and emotion, enjoyment and fun, aesthetics, hedonic quality, engagement etc. The purpose of UX evaluations is to thoroughly understand the user's experiences, be it positive or negative (Vermeeren et al., 2010).

Interviews are also a popular usability/UX method. Interviews provide insights on the views, experiences, beliefs and the motivations of the target users. Research interviews constitute of three types i.e. standardized, semi standardized and unstandardized. Interviews can be conducted, face to face, or telephonic. One can also make use of digital means such as skype or facetime which is a combination of both. Other assessment methods consist of focus group, user observation, analysis of video recordings and diaries etc. Psychophysiological methods such as galvanic skin response, EMG, heart rate etc. are also used as UX assessment tools. Psychophysiological methods are good to assess the emotional reaction of users. Such methods are however rarely used (Bargas-Avila & Hornbæk, 2011). Consequently, we propose the initial requirements RE6.7 – REQ6.13 for C4A system evaluation tools (see Table 12).

Detailed information of user test plans for evaluation of phase 1 demonstrator can be found in D5.1.

Table 12: Initial C4A Evaluation Tools Requirements

Requirement 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	Plans should be identified to conduct user tests for the overall UX of C4A system.
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, hedonic qualities and emotional reactions.
REQ6.7	Data collection tools should be defined in order to collect relevant data at all stages of UX evaluation.
REQ6.8	Various evaluation methods have to be identified for assessing the selected UX components.
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

For satisfying different business models, that are described in detail in D6.3, different system requirements are derived.

The goal of the C4A system is to make TV content accessible for deaf people, satisfying the UN Convention on the Rights of Persons with Disabilities. Therefore, the generated C4A content must be comprehensible for deaf users, best to be applicable to different application areas (like broadcast, government, events, education). Thereby, the C4A system should support operators, primarily broadcasters, in producing sign language content regarding the reference scenario (see section 2.2 and D2.1) by reducing costs and efforts. The following requirements of business models are derived for phase 1 demonstrator.

Table 13: Initial C4A Business Models Requirements

Requirement ID	**Req. description**
REQ7.1	The generated C4A sign language content must be comprehensible for deaf people.
REQ7.2	C4A system architecture is designed so that system can be applied for different business areas.
REQ7.3	With C4A system, sign language TV that is faced by the reference scenario is produced in an easy, rapid and beneficial way (compared to traditional sing language TV production).

4.8. Requirements and Matching C4A objectives

In this section, the identified requirements will be matched to C4A main objectives, listed in the following table.

Table 14: Initial C4A Requirements and Matching to Project Objectives

Objectives	Requirements
1. Sign-interpretation parameter extraction, 3D model generation and real-time animation	1.4, 1.5, 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, 2.14, 2.15
2. Signal processing for sign-language interpretation	1.1, 1.2, 1.3, 1.6, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 3.10, 3.11, 4.1, 4.2, 4.3, 4.4
3. Experimental performance evaluation and creating sustainable business models	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, 7.1, 7.3
4. Building a large-scale and marketable 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, 5.22, 5.23, 5.24, 7.1, 7.2, 7.3

5. Conclusion

This document described the use cases and initial technical, user and evaluation requirements of C4A system phase 1 demonstrator as well as their matching to projects objectives, based upon the general system architecture and reference scenario which are addressed in “D2.1 Initial Reference System Architecture”.

Both, use cases and initial requirements for C4A system, which were derived based upon the four main system components *Broadcaster*, *Remote Studio*, *Processing & Rendering Unit* and *User Terminals*, were identified within this document and mitigation solutions were provided. The use cases and requirements refer to the phase 1 demonstrator, where a sign-interpreter will be captured and inserted into existing TV during the defined reference scenario of News.

The use cases and requirements will be updated along further development, especially for phase 2 demonstrator, during the coming years. The use cases and requirements will be updated for phase 2 demonstrator in D2.4 at M24. Future deliverables covering the details of the architecture, reference scenario, final use cases and requirements as well as planning of C4A demonstrator planning:

- D2.1 “Initial Reference System Architecture” (M6)
- D2.4 “Final Requirements and Specifications for Pilot Demonstrators“ (M24)
- D5.1 “Pilot 1 Demonstrator Architecture, Integration Plan, and Evaluation Methodologies”
- D5.5 “Integrated Pilot 1System and Evaluation Report“ (M24).
- D6.3 “Business Model Analysis” (M6)

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.

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

Brewer, J., Carlson, E., Foliot, J., Freed, G., Hayes, S., Pfeiffer, S., & Sajka, J. (2015). Media Accessibility User Requirements. W3C Working Group Note. Retrieved from http://www.w3.org/TR/media-accessibility-reqs/ [on: 23.01.2018]

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.

Kurz, I., & Mikulasek, B. (2004). Television as a source of information for the deaf and hearing impaired. Captions and sign language on Austrian TV. *Meta: Journal des traducteursMeta:/Translators' Journal*, *49*(1), 81-88.

Kyle, J. G., & Woll, B. (1988). *Sign language: The study of deaf people and their language*. Cambridge University Press.

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.

Liddell, S. K. (2000). Blended spaces and deixis in sign language discourse. In D. McNeill (Ed.) (2000). *Language and gesture*. Vol. 2. Cambridge University Press.

Muir, L. J., & Richardson, I. E. (2005). Perception of sign language and its application to visual communications for deaf people. *Journal of Deaf Studies and Deaf Education*, *10*(4), 390-401.

Orero, P., Serrano, J., Soler, O., Matamala, A., Castella, J., Sanfiel, M. T. S., Vilaro, A., & Mangiron, C. (2014). Accessibility to digital society: Interaction for all. Proce*edings of ICDS 2014. The Eighth International Conference on Digital Society*, 23-27 March 2014, Barcelona.

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.

Wilson, M. & Hoong Sin, C. (2015). Research into the Deaf audience in the UK. A review of evidence. Office for Public Management. Retrieved from http://www.bslzone.co.uk/files/4914/5320/1727/OPM_report_-_Research_into_the_Deaf_Audience_Jan_2016.pdf [on 23.01.2018].

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