Deliverable

Project Acronym:	**IMAC**
Grant Agreement number:	761974
Project Title:	*Immersive Accessibility*

D4.4 - Report on new accessibility formats

Revision: v1.0

Authors: Peter tho Pesch (IRT)

Delivery date: M23 (29-08-2019)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 761974

Dissemination Level		
P	Public	x
C	Confidential, only for members of the consortium and the Commission Services	

Abstract:

This document provides an overview of the signalling and metadata solutions for identifying and describing access services as they are implemented in the ImAc project. Both state-of-the-art signalling means and ImAc extensions are explained.

REVISION HISTORY

Revision	Date	Author	Organisation	Description
0.1	29-07-2019	Peter tho Pesch	IRT	First draft
0.2	05-08-2019	Peter tho Pesch	IRT	Draft shared with partners for feedback
0.3	14-08-2019	Peter tho Pesch	IRT	Content complete except introduction section
0.4	16-08-2019	Peter tho Pesch	IRT	Version for internal review
0.5	23-08-2019	Peter tho Pesch	IRT	Resolved review comments
1.0	29-08-2019	Peter tho Pesch	IRT	Version for submission

EXECUTIVE SUMMARY

In ImAc some new features have been introduced which require metadata in the delivered streams that are not covered by current standards. It is relatively simple to realize such new features in a closed environment like we have initially conducted within the ImAc project. However, one of the projects key objectives is to work towards standardization, and thus interoperable and standard-compliant solutions for providing access services for 360° media applications need to be specified and adopted.

This document is intended to help early adaptors who want to introduce similar services to their portfolio with standard-compliant solutions. We have carefully investigated the needs for integrating additional, i.e. new data, into the content stream(s). Extensions were only made if it wasn't possible to use existing standards.

After evaluating existing standards (mainly MPEG DASH and TTML) and comparing their possibilities with the identified ImAc requirements, we found that the foundations for the different access services are quite diverse. While for the subtitle service, we mostly added 360° relevant data, we needed to find a more extensive solution for the signer service. These are the key findings by each access service:

- Subtitles: Using the IMSC format for subtitles, additional data can be carried within the subtitle format. The adaptation of the subtitle service to a 360° environment requires reconsidering the positioning strategies of subtitles (refer to section 2.4.2) as well as the reference to the spatial position of the related speaker (refer to section 2.4.3). Only one exception: Signalling a subtitle stream as easy-to-read should happen at the MPEG DASH layer, but is currently not specified.
- Audio Description and spoken subtitles: The ImAc specific extensions regarding audio description and spoken subtitles are processed and pre-rendered during the production process on the server side. In the distributed stream, relevant parameters of the various audio streams need to be signalled in order to enable the player application to select the desired stream. We proposed dedicated attributes within the DASH format to identify the different ImAc variations for these services (refer to section 3.3.2 and 4.1). An additional problem is the absence of any standardized solution for identifying an audio stream as Ambisonics (a coding for spatial audio). This is a requirement that we believe others will have as well and thus we have filed a corresponding proposal to MPEG (refer to section 3.3.1).
- Sign Language: Realizing the sign language service was more challenging due to the number of additional information we need and the absence of suitable locations for such data in the media stream. We decided to create a sidecar (accompanying) metadata file carrying all relevant information timed to the media timeline. Embedding the metadata file in the MPEG DASH manifest is described in section 5.3.1. The separate data fields we introduced are described in sections 5.3.2 - 5.3.5.

CONTRIBUTORS

First Name	Last Name	Company	e-Mail
Peter	tho Pesch	IRT	peter.thopesch@irt.de
Ronald	Mies	IRT	ronald.mies@irt.de
Mario	Montagud	i2cat	mario.montagud@i2cat.net

CONTENTS

LIST OF ACRONYMS

Acronym	Description
AAC	Advanced Audio Coding
CAN	Ambisonic Channel Number
AD	Audio Description
AST	Audio Subtitles
CMAF	Common Media Application Format
DASH	Dynamic Adaptive Streaming over HTTP
DVB	Digital Video Broadcast
EBU-TT	EBU Timed Text
EBU-TT-D	EBU Timed Text subtitle distribution format
HMD	Head-mounted Display
HLS	HTTP Live Streaming
IMSC	Internet Media Subtitles and Captions
MPD	Media Presentation Description
MPEG	Motion Picture Experts Group
SL	Sign Language
ST	Subtitles
TTML	Timed Text Markup Language
UI	User Interface
VR	Virtual Reality

1. INTRODUCTION

1.1. Purpose of this document

The purpose of this document is to provide a comprehensive description of the signalling and metadata solutions for identifying and describing access services as they are implemented in the ImAc project. The document should enable other content providers to realize our solutions within their own premises.

In addition to the description of proposed extensions to existing formats, the document comprises a detailed summary of signalling means regarding access services in MPEG DASH.

1.2. Scope of this document

The document's title ("Report on new accessibility formats") suggests a format description restricted to production file formats. Indeed, this was the original intention of D4.4. But preparing and providing all required data from production processes to the user device requires identifying the right location for each piece of information within the different layers of the delivered media stream(s).

The format extensions used for the project's implementation apply not only to the accessibility formats themselves, but also to the streaming format (where we use MPEG DASH in the project). Thus, it was decided to widen the scope of this document. We believe that this decision increases the value of this document to parties that are interested in adopting and/or understanding our solutions.

The document provides a comprehensive overview of all relevant file and streaming formats used in the ImAc project. We consider and describe the possibilities that current state-of-the-art standards provide before adding new file extensions or metadata fields. For each access service type, the reader will find:

- a short recap of the project's aim for this service and identified standard gaps (motivation),
- a summary of the state-of-the-art possibilities in current standards,
- suggested extensions used in ImAc, and
- a summary of the (complete) signalling data and metadata model as used in the project's implementation.

Proprietary mezzanine formats that are used during production by some of the project's tools are not within the scope of this document since they do not contribute to a standardized and open solution for delivering ImAc access services.

1.3. Status of this document

This is version 1.0 of the document, which is also the final version of D4.4. No further updates are planned within the project. The proposed solutions described in this document were found to be valuable extensions that may be used beyond the scope of this project. Although this is the final signalling and metadata model of ImAc, changes of the defined data may happen due to feedback from external parties (e.g. during standardization activities). The ongoing activity of feeding project results into standardization groups will carry on within task T6.2.

1.4. Relation with other ImAc activities

This document summarizes technical solutions for providing data that were investigated and discussed during design and implementation tasks throughout the whole ImAc project. As such, the document relates to many activities:

- The production tools (T4.1, T4.2, T4.3), the ACM (T3.2), the packager (T3.3) as well as the player (T3.5). In these tasks the signalling data and metadata that are summarized in this document have been discussed and defined.
- In T3.4, the signalling of access services is related to the user interface of the player. The deliverable D3.4 [1] provides an overview of ImAc features not supported in state-of-the-art distribution standards today (gap analysis).
- T3.6 and its complemental deliverable D3.6 [2], which is dedicated to the integration of access services in an end-to-end workflow, relate to this document as well.
- Some of the proposed solutions described in this document have resulted in standardization activities, which are carried out in T6.2.

2. SUBTITLES

2.1. Motivation

In the beginning of the ImAc project, we have gathered a list of user requirements for presenting subtitles in 360° videos. Based on this list, we have evolved several presentation options, derived from user-centric activitites, and evaluated through various test phases (described in detail in D5.2 [3]). We found that current standards and specifications do not provide the required possibilities to realize all features of the developed presentation options. In this document, we describe possible solutions in detail, based on existing standards.

In particular, we have identified the following requirements that arise from the ImAc use cases[1]:

1. Means to signal a subtitle service' type as "easy to read"
2. Means to position subtitle regions on a 2D plane in the field of view of the user.
3. Means to locate the audio source of the timed text horizontally (e.g. with a longitude coordinate).

Note: *Deliverable D3.4, chapter 5.1, provides a gap analysis regarding the signalling of ImAc services and an overview of the required ImAc extensions*.

Sections 2.2 and 2.3 describe relevant data for subtitles in the state-of-the-art standard MPEG DASH and IMSC specifications.

Proposals for possible solutions to meet these requirements, based on DASH and IMSC, are described in Section 2.4.

Section 2.5 provides a summary of relevant signalling data, as is used in the ImAc implementation using existing standard means as well as ImAc specific extensions. In some parts, the ImAc implementation does not fully follow the suggested solutions described in Section 2.4, e.g. when we propose to extend existing (standard) descriptors with additional values we will use our own user defined descriptor in order to be standard compliant with the current implementation.

In addition, a presentation option has been tested in the pre-pilots2 tests that is not covered by above-standing requirements. In this presentation option, subtitles were displayed next to the speaker, fixed to the 360° video. The need for such a presentation mode has not been confirmed by the conducted tests and for that reason this presentation option was not added to the ImAc player and will not be followed during the project's second pilot phase. To follow up on this option, more investigations are needed, on the way the presentation option works in detail as well as on the technical realization.

2.2. State of the art – TTML subtitles in DASH

IMSC subtitles can be signalled and described in MPEG DASH [4] using the DVB-DASH profile [5] as summarized in the following sub-chapters.

[1] The use cases that were tested with end users during various test phases are summarized in D5.2, second iteration [3]. Features for subtitles are listed in chapter 3.1.1.1, features for easy-to-read subtitles are listed in chapter 3.1.1.2 of this document.

2.2.1. `Role` element

Location:
`AdaptationSet/ContentComponent` elements - as part of an MPD

Defined in:
- ISO/IEC 23009-1:2014 [4], § 5.8.4.2, § 5.8.5.5
- ETSI TS 103 285 V1.1.1 [5], § 7.1.2

Description:

The Role element is an element in DASH to specify an arbitrary role for the affected component that is based on a specific scheme. The element can be used multiple times on a component.

`Role` element – `@schemeIdUri` attribute
This attribute specifies the scheme used for the role value. The DASH specification itself defines a simple role scheme (URI: `urn:mpeg:dash:role:2011`) which shall be used here.

`Role` element – `@value` attribute

This attribute specifies the actual role value. Regarding subtitles, the applicable values depend on the actual purpose of the subtitles. The following values from the mentioned scheme are used here:

- main
- alternate
- commentary

2.2.2. Accessibility element

Location:
`AdaptationSet/ContentComponent` elements - as part of a MPD

Defined in:
- ISO/IEC 23009-1:2014 [4], § 5.8.4.3
- ETSI TS 103 285 V1.1.1 [5], § 7.1.2
- ETSI TS 102 822-3-1 V1.9.2 [6], § A.15

Description:

The Accessibility element is a generic element in DASH to specify an arbitrary accessibility property that is based on a specific scheme. The element can be used multiple times on a component. Regarding subtitles this element is only used for subtitles for the hard of hearing.

Accessibility element – @schemeIdUri attribute
This attribute specifies the scheme used for the accessibility value. DASH itself doesn't define a scheme here. For subtitles for the hard of hearing, the *Audio Purpose* scheme from the *TV Anytime* project is used (URI: urn:tva:metadata:cs:AudioPurposeCS:2007).

Note: It is indeed correct that a scheme actually designated for audio is used for subtitles here.

Accessibility element – @value attribute

This attribute specifies the actual accessibility value. To indicate subtitles for the hard of hearing, the value `2` from the mentioned scheme is used here.

2.2.3. AdaptationSet element

@mimeType attribute

Location:

`AdaptationSet/Representation/SubRepresentation` elements - as part of an MPD

Defined in:

- ISO/IEC 23009-1:2014 [4], § 5.3.7.2
- ETSI TS 103 285 V1.1.1 [5], § 7.1.1

Description:

The mandatory mimeType attribute is used to describe the MIME type of the actual media segments.

For TTML subtitles, the value "application/mp4" (TTML encapsulated in ISOBMFF) is used.

`@codecs` attribute

Location:

`AdaptationSet/Representation/SubRepresentation` elements - as part of an MPD

Defined in:

- ISO/IEC 23009-1:2014 [4], § 5.3.7.2
- ETSI TS 103 285 V1.1.1 [5], § 7.1.1

Description:

The optional codecs attribute specifies the used codec. It shall also include profile and level information, if applicable. For TTML subtitles, the value shall begin with "stpp" indicating XML subtitles. The value should indicate the used profile according to the W3C TTML profile registry. If the value "stpp" is used, the value "stpp.ttml.etd1" shall be assumed, referring to TTML content suitable for an EBU-TT-D renderer.

`@lang` attribute

Location:

`AdaptationSet/ContentComponent/ProgramInformation` elements - as part of an MPD

Defined in:

- ISO/IEC 23009-1:2014 [4], § 5.3.3.2
- ETSI TS 103 285 V1.1.1 [5], § 7.1.2

Description:

The lang attribute is used to describe the actual language of e.g. a subtitle adaptation set. Its value is specified according to RFC 5646 [18].

2.3. State of the art – positioning of subtitles in IMSC

The following description of positioning subtitles within the IMSC root container assumes the writing direction left to right and top to bottom. Note that the writing direction of a subtitle language may affect the positioning of subtitles within a region.

2.3.1. Root container region

Location

n.a.

Defined in

- W3C IMSC 1.0.1 [7], §6.7.1
- W3C TTML1 [8], §2.2

Description

The Root Container Region of a Document Instance is mapped to each image frame of the Related 2D Video Object. The presence of attribute aspectRatio may result into a root container region that doesn't match the image frame in its entirely.

2.3.2. Region element

Location

/tt/head/layout

Defined in

- W3C IMSC 1.0.1 [7], §6.8.2
- W3C TTML1 [8], §9.1.2

Description

Defines a space or area for the display of subtitle content. The region shall not extent the root container region.

@origin attribute

The origin attribute sets the offset of the region of the top left corner of the root container region.

@extent attribute

The extent attribute sets the width and height of the region.

2.3.3. Style element

Location

/tt/head/styling

Defined in

- W3C IMSC 1.0.1 [7], §7.4
- W3C TTML1 [8], §8.1.2

Description

The styling element contains a set of style information through style attributes. Among others, the (horizontal) alignment of subtitle blocks is defined via the tts:textAlign attribute.

@textAlign attribute

It defines the alignment of a subtitle block. Allowed values are “left”, “center”, “right”, “start”, “end”

2.4. Proposed solutions for ImAc services

As described in Section 2.1, we identified the following features that are not sufficiently supported by relevant standards:

1. Means to signal a subtitle service type as “easy-to-read”
2. Means to position subtitle regions on a 2D plane in the field of view of the user.
3. Means to locate the audio source of the timed text horizontally (e.g. with a longitude coordinate).

The following sub-sections describe possible solutions to realize these features. The ImAc implementation is described separately from the proposed solution because we did not always implement the suggested solutions entirely to avoid standard conflicts. For example, when we propose to extend existing (standard) descriptors with additional values, in the current implementation we will use our own user defined descriptor in order to be standard compliant with the current implementation.

2.4.1. Easy to read subtitles

Easy-to-read subtitles are an alternative subtitle service / track that may be present in addition to the standard subtitle track (in the same language). Alternatively, the (only) subtitle track for a language may meet the conditions of "easy-to-read" and may be labelled accordingly.

Regarding the meaning of easy-to-read, the ImAc project refers to the description in ITU-R BT.2207-3 [9], chapter 6. In short, easy-to-read subtitles is a broadcast service tailored to the language comprehensive level of non-native speakers, children, people with developmental impairments and people with cognitive losses due e.g. to aging.

The signalling for easy-to-read must be available to the player, ideally without the need to decode each available subtitle track. Typically, a player would show the available subtitle services on a User Interface (UI) to the users, such that they can select the preferred option.

Proposed solution

Following this requirement, we propose to signal easy-to-read on the DASH layer (or equivalent when using other streaming formats, like HLS or CMAF). This value shall identify the labelled media as intended for people who need a simple language to be able to follow the content. The labelled media should follow the recommendations for easy-to-read language described in ITU-R BT.2207-3 [9], chapter 6.

For the usage of MPEG DASH, we propose to add and easy-to-read identifier as a new value to the descriptor "urn:tva:metadata:cs:AudioPurposeCS:2007" for the Accessibility element. This descriptor uses integer values that are mapped to target audience groups. For example, the value "2" is used to indicate the target group "for the hard of hearing". The format of such a new value must still be discussed.

An alternative solution would be to signal easy-to-read in the subtitle format, in our case IMSC. But a player would need to decode all subtitle files in order to determine if a track contains easy-to-read subtitles. Although this is a simple readout, it would be an unnecessary task in the player.

Additional proposal for the TTML file format

To prepare the subtitle file for the correct signalling in distribution, the subtitle file should be labelled accordingly during the authoring stage. One possibility would be to add this as an additional specific content type for "Easy to Read" in the EBU-TT Classification Scheme [10] used in the ebuttm:documentContentType element, which is used in EBU-TT to set the content type for certain subtitle content.

Sample implementation in ImAc:

The ImAc implementation defines a custom descriptor for the accessibility element, such that no conflicts with the existing standards occur. It is intended to keep this custom solution in place until a standardized solution becomes available. Accordingly, the signalling of easy-to-read subtitles in ImAc is as follows:

- Accessibility element with schemeIdUri "urn:imac:access-identifier:2019" and value "easy-to-read".

MPD sample snippet (ImAc implementation):

Next, an example of an MPD snippet including the signalling of easy-to-read subtitles, as implemented in ImAc, is provided.

```
<AdaptationSet>

    <Role
        schemeIdUri="urn:mpeg:dash:role:2011"
        value="alternate"/>

   <Accessibility
        schemeIdUri="urn:imac:access-identifier:2019"
        value="easy-to-read"/>
    …


</AdaptationSet>
```

2.4.2. Positioning of a subtitle region in 360° scenes

The positioning means of IMSC currently do not include the use case of a 360° video environment and consequently do not allow defining the position of the subtitle region in a 360° video or even a 3D scene.

Two general approaches for subtitle positioning were tested in the ImAc project:

1. Rendering subtitles always visible fixed to the screen / viewport (i.e. field of view)
2. Rendering subtitles fixed to the video (e.g. attached to the speaker)

Our proposal refers to the first approach (subtitles always visible fixed to screen). As described in Section 2.1, no proposal for subtitles fixed to the video is made at this stage, since a corresponding user requirement could not be confirmed by the project's user tests (for user test results refer to D5.4 [11]).

The IMSC specification (W3C IMSC 1.0.1 [7]) requires the allocation of a "root container region" that establishes a 2D coordinate system where regions can be rendered into. IMSC further specifies that the root container region spans over the entire image frame (special rules apply if the attribute ittp:aspectRatio is set though).

Following the TTML specification, the properties of the root container region may be established by the document processing context. But such a processing context has yet to be defined for a 360° scene.

We found that neither TTML itself, nor any of the TTML-derived formats (like IMSC), provides a 2D coordinate system within the 360° scene that may be referred to as the root container region.

Proposed solution

We will describe a possible mapping of a 2D coordinate system into a 360° scene. In this context, we understand the 360° scene as a 3D space with the following constraints:

- The 3D space's purpose is to render an image for the presentation on a screen. The screen may be a tablet or PC screen, or the (stereoscopic) screen of a Head Mounted Display (HMD).

- The image is retrieved by a rendering process that calculates the picture that a virtual camera would capture. The virtual camera is located at the origin of the coordinate system that spans the 3D space. The camera may be stereoscopic. The rotation of the virtual camera and its optical parameters define the part of the scene that is visual on the retrieved image.
- The 3D space further contains a non-stereoscopic video mapped onto a geometrical shape which center of mass lies at the origin of the coordinate system that spans the 3D space. Typically, the geometrical shape is a sphere or a cube, depending on the encoding of the 360° video image.

In the 360° scene with these properties, we suggest spanning a 2D coordinate system that represents the root container region of an IMSC document. The plane spanned by the 2D coordinate system with the axis X and Y shall have the following properties:

- The plane shall lie between the origin of the 3D coordinate system and the video's geometrical shape but may intersect with this shape.
- The plane shall lie orthogonal to the vector that goes through the origin of the 3D coordinate system and the center of the current viewport. The point where this vector intersects the 2D plane is also the origin of the 2D coordinate system.
- One of the axis of the 2D plane shall be parallel to the horizontal plane spanned by the 3D coordinate system.

The positioning of the subtitle within the root container region shall largely follow the IMSC specification, i.e. by processing the corresponding layout (parameter) attributes. However, we further suggest tailoring the positioning of subtitles to the target device as described below.

When rendering onto a 2D screen (e.g. when watching the 360° video on a tablet or PC monitor):

- The IMSC subtitle plane (root container region) should be mapped to the field of view / viewport. That means the root container region spans over the entire viewport. If the attribute "ittp:aspectRatio" is present, the root container region shall be mapped onto the viewport in the same way as it would be mapped onto an image frame following the description in the IMSC specification, section 6.7.1.
- The IMSC attributes tts:displayAlign, tts:origin, tts:extent on the tt:region element should be processed according to the IMSC specification.
- Text alignment should always be set to "center", the tt:style attribute tts:textAlign should be ignored.

When rendering in VR-mode (e.g. in a HMD):

- The positioning within the field of view should not relate to the region attributes tts:origin and tts:extent.
- Instead, the vertical positioning of the subtitle should be set as a length value (metric: percentage) that specifies how far the subtitle is set from the bottom or top of the viewport. A new attribute should be added to the IMSC format to carry this information. The attribute is considered to be significant only when specified on a tt:region element.
- A new attribute should be added to the IMSC format that defines if subtitles are top aligned or bottom aligned when in VR-mode. The attribute is considered to be significant only when specified on a tt:region element.
- Text alignment should always be set to "center", the tt:style attribute tts:textAlign should be ignored.

- In the stereoscopic rendering, the depth of the subtitle plane within the 360° scene affects the perceived image on the HMD. The distance of the subtitle plane from the virtual camera is left to the player implementation. However, it is suggested, to put the subtitle plane close to the geometrical shape the video is projected on (otherwise, e.g. vision focus issues may appear).

Sample implementation in ImAc:

The current ImAc implementation differs from the above standing suggestion, since the player does not process all style attributes of the IMSC file. The position of subtitles in the viewport is a player internal implementation with the following aspects:

- The player defines three possible rendering areas for subtitles of different sizes where the user can choose from. The format and size of the rendering areas are determined by users' studies targeted at identifying appropriate values for "safe areas" (aka. "comfortable field of view"). All areas are positioned in the center of the viewport and define the margin of the subtitles from the bottom or top of the viewport. The user can select one of the three rendering areas to his / her wishes.
- The player provides the user option to set subtitles top aligned or bottom aligned.
- The player always put subtitles horizontally centered in the viewport.
- Layout data from the subtitle stream is not processed by the player, but is rather considered as a potentially useful option for the subtitle author to describe a preferred vertical position of the subtitle.

2.4.3. Location of audio source

Short summary of ImAc's subtitle presentation options

Information on the (horizontal) location of the audio source is required to realize the enhanced subtitle presentation options developed in the project. The enhancement comprises two different ways of indicating the speakers position in the 360° scene for the active subtitle. The indication is done in form of a visual cue for which two options have been developed:

a) Arrow indicator – An arrow is placed left or right next to the subtitle, pointing into the direction of the related speaker, if the speaker is currently not visible within the picture. The arrow disappears as soon as the speaker becomes visible (i.e. moves into the picture) and appears as soon as the speaker disappears.
b) Radar indicator – a radar system is rendered next to the subtitle (e.g. on the right side) showing the current viewing direction and viewport. The radar is always active. For each active subtitle, a colored dot matching the subtitle color is rendered into the radar system indicating the horizontal position of the related speaker within the 360° scene.

Proposed solution

The IMSC format uses regions and layout attribute to position subtitles within a dedicated rendering plane (the root container region). Using these possibilities, subtitles can be placed next to a person, however subtitles are usually not authored that way. But a direct indication of a speaker position within the picture is supported by neither IMSC nor TTML.

The subtitle presentation options presented by the ImAc project require the information about the (horizontal) position of the related speaker in order to render a visual cue in form of an arrow or a radar system as described above.

The information is required for each timed text segment and thus can be transported on the subtitle format level. This is the simplest way to identify and describe separate timed text (subtitle) segments.

We propose to add an attribute that describes the horizontal direction of an audio source. The attribute should be considered to be significant only when it is applied to a tt:p or a tt:span element. The information may be used by the player application to provide graphical or other indicators for the position of the speaker in the scene.

Sample implementation in ImAc:

An attribute that describes the horizontal direction of an audio source was added under a custom namespace. The attribute is processed only on a tt:p element and has the following properties:

- Namespace used: http://www.imac-project.eu
- Abbreviation used: imac
- Attribute name: equirectangularLongitude
- Value range: [-180, 180]

The center point [0; 0] is defined as the point on the sphere where the center of the equirectangular video texture is projected. Turning left results in increasing (positive) longitude values. Turning right results in decreasing (negative) longitude values.

Sample IMSC snippet (ImAc implementation):

```
<tt:p
    xml:id="s1"
    region="bottom"
    style="defaultStyle"
    begin="00:00:01.000"
    end="00:00:04.000"
    imac:equirectangularLongitude="-30">

    <tt:span
        style="colorYellow">

        Sample subtitle

    </tt:span>

</tt:p>
```

2.5. Summary of signalling used in ImAc

The following tables provides an overview of the transportation of subtitle content and metadata as used in ImAc. That includes both standardized signalling means as well as ImAc extensions. Styling related annotation is not listed separately (e.g. font color, font family,

background color). However, attributes and elements dedicated to subtitle positioning are listed, since this is one of the main aspects of the ImAc investigations.

Some annotation attributes are overwritten by the player implementation, either to adapt subtitles to the 360° viewing environment, or due to personalization option.

> ***Note:*** *There are minor deviations between the described metadata in the tables below and the description in Annex II of deliverable D3.5 [12]. D3.5 describes the metadata as implemented for the pre-pilot 2 tests. After these tests, the signalling and descriptive data for subtitles has been refined to be compatible with state-of-the-art standards. The latest status is listed in this document.*

The meaning of most fields in the table are self-explaining, but few notes are listed below for a richer comprehension:

- The field "Location" describes from where the ImAc decoder (i.e. the ImAc player) retrieves the information. That means in what layer the information is located.
- The field "ImAc extension" indicates if the data is transported using standardized formats of if a format extension that was defined in the ImAc project is used.
- The numbering does not have any specific meaning and is just used for indexing the listed elements.

<table>
<tr><td># 1</td><td><u>Title</u>
Subtitle presence</td><td><u>Location</u>
MPEG DASH manifest</td></tr>
<tr><td colspan="3"><u>Description</u>
Indication, if/where one or more subtitle tracks are available for a program.</td></tr>
<tr><td colspan="3"><u>Usage in ImAc / Explanation</u>
For each subtitle track an AdaptationSet element is added including the track's metadata and source file URL.</td></tr>
<tr><td colspan="2"><u>ImAc extension</u>
NO</td><td><u>Relevant standards</u>
MPEG DASH (ISO/IEC 23009-1:2014)
DVB DASH Profile (ETSI TS 103 285 V1.1.1)</td></tr>
</table>

Sample

```
<AdaptationSet
    contentType="text"
    mimeType="application/ttml+xml"
    segmentAlignment="true"
    lang="eng">

    <Role
        schemeIdUri="urn:mpeg:dash:role:2011"
        value="main"/>

    <Accessibility
        schemeIdUri="urn:tva:metadata:cs:AudioPurposeCS:2007"
        value="2"/>

    <Representation
        id="xml_eng"
        bandwidth="1000">

        <BaseURL>sub_eng_short.xml</BaseURL>

    </Representation>

</AdaptationSet>
```

# 2	**Title** Language	**Location** MPEG DASH manifest and IMSC
Description The language of the subtitles		
Usage in ImAc / Explanation The language can be set in the IMSC file as well as in the DASH manifest. The ImAc player uses the language identified in DASH. The language attribute is also indicated for other tracks, like main audio, audio description and even sign language.		
ImAc extension NO	**Relevant standards** MPEG DASH (ISO/IEC 23009-1:2014) [4] DVB DASH Profile (ETSI TS 103 285 V1.1.1) [5]	

Sample

```
<AdaptationSet
    contentType="text"
    lang="eng" … >
    ...
</AdaptationSet>
```

<table>
<tr><td># 3</td><td><u>Title</u>
Target role</td><td><u>Location</u>
MPEG DASH manifest</td></tr>
<tr><td colspan="3"><u>Description</u>
Indicate the target role of the subtitle stream, i.e. it’s intended purpose. For instance, if the subtitles are “heard of hearing”, "translation" or “easy to read”.</td></tr>
<tr><td colspan="3"><u>Usage in ImAc / Explanation</u>
Subtitles for the hard of hearing are signalled as follows:
• Role element with schemeIdUri “urn:mpeg:dash:role:2011” and value “main”
• Accessibility element with schemeIdUri “urn:tva:metadata:cs:AudioPurposeCS:2007” and value “2”

Subtitles for translation are signalled as follows:
• Role with schemeIdUri “urn:mpeg:dash:role:2011” and value “alternate”
• no Accessibility element present

Easy to read subtitles are signalled as follows:
• Role with schemeIdUri “urn:mpeg:dash:role:2011” and value “alternate”
• Accessibility element with schemeIdUri “urn:imac:access-identifier:2019” and value “easy-to-read”</td></tr>
<tr><td colspan="2"><u>ImAc extension</u>
YES</td><td><u>Relevant standards</u>
MPEG DASH (ISO/IEC 23009-1:2014) [4]
DVB DASH Profile (ETSI TS 103 285 V1.1.1) [5]</td></tr>
<tr><td colspan="3"><u>Sample</u>
<pre>
<AdaptationSet>

    <Role
        schemeIdUri="urn:mpeg:dash:role:2011"
        value="main"/>

   <Accessibility
        schemeIdUri="urn:tva:metadata:cs:AudioPurposeCS:2007"
        value="2"/>
    …

</AdaptationSet>
</pre></td></tr>
</table>

# 4	**Title** Subtitle styling	**Location** IMSC and ImAc Player
Description Styling and alignment of subtitle text, e.g. font attributes, alignment, foreground and background color.		
Usage in ImAc / Explanation The IMSC format includes various attributes to define text layout. ImAc uses these features. However, some attributes are overwritten by the player, either because they are subject of user settings, or to cope with the specific environment of a Head Mounted Display. The main elements and attributes that are overwritten by the ImAc player are: • tts:fontFamily • tts:fontSize • tts:textAlign		
ImAc extension NO	**Relevant standards** W3C IMSC1.0.1 (TTML Profiles for Internet Media Subtitles and Captions v1.0.1) [7]	

# 5	**Title** Region (safe area)	**Location** ImAc Player
Description Defines the part of the field of view where subtitles may be rendered.		
Usage in ImAc / Explanation The IMSC format defines a region element that is used to position subtitles on the screen in 2D. In the ImAc player, the region element is ignored and the positioning of subtitles on screen are handled by the ImAc player and the user settings in order to cope with the specific viewing environment in a Head Mounted Display. The values of the “safe area” in the player are determined by the results of users’ tests conducted to find out the values that lead to a more comfortable viewing experience.		
ImAc extension NO	**Relevant standards** –	

# 6	**Title** Vertical position	**Transported in** ImAc Player
Description The vertical subtitle position within a region.		
Usage in ImAc / Explanation The IMSC format defines the attribute “tts:displayAlign” for vertical alignment of a subtitle within a region on a 2D screen In the ImAc player, the tts:displayAlign attribute is ignored. Instead, the vertical alignment is controlled via user settings, and positioned to fit the limits of the active “safe area”.		
ImAc extension NO	**Relevant standards** –	

# 7	**Title** Presentation mode	**Transported in** ImAc Player
Description Differs between presentation modes provided by the player. Currently supported ones are: • Subtitles with a guiding mechanism (arrow, radar, none) • Auto-positioning (only in test version) • Fixed to video (only in test version)		
Usage in ImAc / Explanation The presentation modes use the annotation that is described in #8 “360° position”. The behaviour for the different modes is an internal player development and does not follow any annotation data from the delivered content stream(s) and metadata files. It has been discussed if the subtitle author should have the possibility to indicate (in the subtitle file) which of the presentation modes will suit the content. This would require an additional metadata element or attribute. However, at the time this document was created, this functionality was not implemented or planned. (Refer to D3.5, chapter 7.5, for detailed information on the different presentation modes)		
ImAc extension NO	**Relevant standards** –	

<table>
<tr><td># 8</td><td><u>Title</u>
360° position</td><td><u>Transported in</u>
IMSC (ImAc extension)</td></tr>
<tr><td colspan="3"><u>Description</u>
Horizontal position of subtitle’s speaker (or an audio source) in the 360° scene. Provided as longitude value.</td></tr>
<tr><td colspan="3"><u>Usage in ImAc / Explanation</u>
The attribute in a user defined namespace is set on tt:p elements.
Longitude value rage: [-180; 180]

Namespace declination: xmlns:imac="http://www.imac-project.eu"

The center point [0; 0] is defined as the point on the sphere where the center of the equirectangular video texture is projected. Viewing directions up and left result in increasing latitude and longitude values.</td></tr>
<tr><td colspan="2"><u>ImAc extension</u>
YES</td><td><u>Relevant standards</u>
W3C IMSC1.0.1 (TTML Profiles for Internet Media Subtitles and Captions v1.0.1) [7]</td></tr>
<tr><td colspan="3"><u>Sample</u>
<pre>
<tt:p
    xml:id="s1"
    region="bottom"
    style="defaultStyle"
    begin="00:00:01.000"
    end="00:00:04.000"
    imac:equirectangularLongitude="-30">

    <tt:span
        style="colorYellow">

        Sample subtitle

    </tt:span>

</tt:p>
</pre></td></tr>
</table>

3. AUDIO DESCRIPTION

3.1. Motivation

In Deliverable D3.4, chapter 5.1 [1], we provided a gap analysis regarding the signalling of ImAc services and an overview of the required ImAc extensions. In this document, we describe possible solutions in detail, based on existing standards. Section 3.2 describes relevant data for audio and audio description in particular in the state-of-the-art MPEG DASH specification.

The requirements for audio description (AD) that arise from the ImAc services are:

1. Means to signalize an Ambisonics audio stream in the delivery (streaming) format MPEG DASH.
2. Means to signalize custom properties of audio streams in the delivery (streaming) format MPEG DASH.

Proposals for possible solutions to meet these requirements, based on DASH, are described in Section 3.3.

3.2. State of the art

Audio Description can be signalled and described in MPEG DASH using the DVB-DASH profile as summarized in the following sub-chapters.

3.2.1. `Role` element

Location:
`AdaptationSet`/ContentComponent elements - as part of an MPD

Defined in:
- ISO/IEC 23009-1:2014 [4], § 5.8.4.2, § 5.8.5.5
- ETSI TS 103 285 V1.1.1 [5], § 6.1.2

Description:

The Role element is an element in DASH to specify an arbitrary role for the affected component that is based on a specific scheme. The element can be used multiple times on a component.

Role element – @schemeIdUri attribute
This attribute specifies the scheme used for the role value. The DASH specification itself defines a simple role scheme (URI: `urn:mpeg:dash:role:2011`) which shall be used here.

`Role` element – `@value` attribute

This attribute specifies the actual role value. Regarding audio description, the applicable values depend on the actual purpose of the audio description. The following values from the mentioned scheme are used here:

- alternate (indicating Broadcast mix AD)

- commentary (indicating Receiver mix AD)

3.2.2. Accessibility element

Location:
`AdaptationSet/ContentComponent` elements - as part of a MPD

Defined in:
- ISO/IEC 23009-1:2014 [4], § 5.8.4.3
- ETSI TS 103 285 V1.1.1 [5], § 6.1.2
- ETSI TS 102 822-3-1 V1.9.2 [6], § A.15

Description:

The Accessibility element is a generic element in DASH to specify an arbitrary accessibility property that is based on a specific scheme. The element can be used multiple times on a component.

Accessibility element – @schemeIdUri attribute

This attribute specifies the scheme used for the accessibility value. DASH itself doesn't define a scheme here. For audio description, the *Audio Purpose* scheme from the *TV Anytime* project is used (URI:urn:tva:metadata:cs:AudioPurposeCS:2007).

Accessibility element – @value attribute

This attribute specifies the actual accessibility value.

To indicate audio description, the value `1` from the mentioned scheme is used here.

3.2.3. SubRepresentation element

@mimeType attribute

Location:
`AdaptationSet/Representation/SubRepresentation` elements - as part of an MPD

Defined in:
- ISO/IEC 23009-1:2014 [4], § 5.3.7.2
- ETSI TS 103 285 V1.1.1 [5], § 6.1.1

Description:

The mandatory mimeType attribute is used to describe the MIME type of the actual media segments.

For audio description, the value "audio/mp4" is used for all audio codecs mentioned in the audio chapter (ETSI TS 103 285 V1.1.1, §6).

`@codecs` attribute

Location:
`AdaptationSet/Representation/SubRepresentation` elements - as part of an MPD

Defined in:
- ISO/IEC 23009-1:2014 [4], § 5.3.7.2
- ETSI TS 103 285 V1.1.1 [5], § 6.1.1

Description:

The mandatory codecs attribute specifies the used codec. It shall also include profile and level information, if applicable. For audio description, the value depends on the audio codec actually used and starts with a matching codec identifier.

3.2.4. ProgramInformation element

`@lang` attribute

Location:
`AdaptationSet/ContentComponent/ProgramInformation` elements - as part of an MPD

Defined in:
- ISO/IEC 23009-1:2014 [4], § 5.3.3.2
- ETSI TS 103 285 V1.1.1 [5], § 6.1.2

Description:

The lang attribute is used to describe the actual language of e.g. an audio description adaptation set. Its value is specified according to RFC 5646.

3.2.5. Representation element

@dependencyId attribute

Location:
`Representation` element - as part of a MPD

Defined in:
- ISO/IEC 23009-1:2014 [4], § 5.3.5.2
- ETSI TS 103 285 V1.1.1 [5], § 6.1.2

Description:

The dependencyId attribute, if present, indicates that a representation is a dependent representation and hence depends on other representations in order to be rendered. The attribute value consists of a space-separated list of one or more representations (actually: their IDs) this representation depends on.

Regarding audio description this element is only used for receiver-mix AD.

3.3. Proposed solutions for ImAc services

The ImAc project identified the following features that are not sufficiently supported by relevant standards.

- Signalling of Ambisonics
- Description of audio content variations

The following sub-sections describe possible solutions to realize these features. The proposals contain both standard extensions and the specific solutions used in ImAc.

3.3.1. Signalling of Ambisonics in MPEG DASH

For ImAc's spatial audio services, audio will be delivered as Ambisonic audio (1st order) using generic audio compression formats, in our case AAC. This may be a common use case, since Ambisonic decoder libraries are available and there are only few alternatives currently available to realize spatial audio on a wide range of client devices. We realized that MPEG-DASH is currently missing the possibility of signalling this at the MPD level, such that a player would be allowed to pick the right set of audio streams according to its capabilities.

Another use case is the possibility to have at the same time an Ambisonic representation and a mono/stereo non spatialized track (usually called head-locked). This sound data could be carried in a dedicated representation or embedded within the stream containing the Ambisonic data. For the latter case, an extension of the channel mapping would be needed to signal either a mono channel or a stereo pair (left / right).

Proposal

We suggest introducing a new descriptor for Ambisonic covering our use case. We propose to define an AmbisonicChannelMap descriptor with the following properties:

urn: `urn:mpeg:dash:ambi-map:2018`

value syntax: a comma separated list of either integer or 'L','R' or 'M'

Integers specify the Ambisonic Channel Number (ACN), 'M' indicates a mono channel, 'L' a left and 'R' a right channel.

semantics: The value field gives the mapping between the channels present in the stream and their corresponding ACN or mono/left/right ordered list of channels. 'M','L','R' indicate "head-lock" channels which shall not be spatialized but still mixed in the output.

position: when present, this descriptor shall be either at the AdaptationSet or Representation level in the MED

optionality: this descriptor may be signaled as a Supplemental property if the stream contains only the first order Ambisonic channel, or the mono head-lock or the full stereo pair head-lock. Otherwise, it shall be signaled as an Essential property.

Sample integration in ImAc

The proposed solution has not been implemented in ImAc yet. It is intended to switch to the proposed descriptor for Ambisonic signalling.

MPD snipped for proposed solution

```
<AdaptationSet …>

    <EssentialProperty
        schemeIdUri="urn:mpeg:dash:ambi-map:2018"
        value="0 1 2 3"/>

        <Representation …>
            ... // 4 channels audio
        </Representation>

</AdaptationSet>

<AdaptationSet>
    //stereo, option 1

    <Representation>
        <AudioChannelConfiguration
            schemeIdUri="urn:mpeg:dash:23003:3:audio_channel_configuration:2011"
            value="stereo"/>
        ... // 2 channels audio
    </Representation>

</AdaptationSet>

<AdaptationSet>
    //stereo, option 2

    <SupplementalProperty
        schemeIdUri="urn:mpeg:dash:ambi-map:2018"
        value="L R"/>

        <Representation …>
            ... // 2 channels audio
        </Representation>

</AdaptationSet>
```

3.3.2. Audio variations

The ImAc audio services rely on the delivery of several audio streams for each service. A service in this context is for example Audio Description mixed with the main audio in one language. Each stream contains a different audio mix and/or different audio treatment. We must signal the different treatments to the player application such that it can pick the intended stream for playback.

ImAc has introduced two parameters for the project's purposes:

- AD Gain – The gain of the AD part of the signal in the mix (when using broadcast-mix mode)

- Audio mode – Spatial audio treatment, e.g. used to spatially position AD segments in the (audio) scene. [2]

The use case to identify custom parameters on the level of the streaming format (MPEG DASH in our case) is not restricted to this project, but a general need when introducing features as the above mentioned.

Proposal

We propose the use of a custom namespace attribute to carry the necessary information.

Sample integration in ImAc

For the ImAc implementation, two custom attributes were added to the DASH manifest. A separate namespace was defined to avoid conflicts in player applications. The attributes are added to the Representation element under the namespace "http://www.imac-project.eu" with the abbreviation "imac".

@audioDescriptionGain
Values: "low", "medium", "high"

@audioMode
Values: "classic", "static", "dynamic"

Sample MPD snippet (ImAc implementation):

```
<AdaptationSet>

    <Representation
        imac:audioDescriptionGain="medium"
        imac:audioMode="dynamic">

        ...

    </Representation>

    <Representation
        imac:audioDescriptionGain="medium"
        imac:audioMode="classic">

        ...

    </Representation>

</AdaptationSet>
```

[2] The audio modes are explained in Deliverable D5.2, second iteration in chapter 3.1.1.4 [3]

3.4. Summary of signalling used in ImAc

The following tables provide an overview of the transportation of Audio Description signalling and metadata as used in ImAc.

The meaning of most fields in the table are self-explaining, but few notes are listed below for a richer comprehension:

- The field “Location” describes from where the ImAc decoder (i.e. the ImAc player) retrieves the information. That means in what layer the information is located.
- The field “ImAc extension” indicates if the data is transported using standardized formats of if a format extension defined in the ImAc project is used.
- The numbering does not have any specific meaning and is just used for indexing the listed elements.

<table>
<tr><td rowspan="2"># 1</td><td><u>Title</u>
AD presence and target group</td><td><u>Location</u>
MPEG DASH manifest</td></tr>
<tr><td colspan="2"><u>Description</u>
Indication, if and where Audio Description is available in the stream.</td></tr>
<tr><td colspan="3"><u>Usage in ImAc / Explanation</u>
The presence of AD within an audio track is signalized by the combination of two elements.

Accessibility:
Set to "1" for tracks with AD. Otherwise element is absent.

Role:
The value "alternate" is suggested to use for Broadcast mix AD (this is how it's done in ImAc). Additional values are 'commentary' which is used for Receiver mix AD. The value 'alternate' is also used for clean audio.

The value 'main' is set for the audio tracks without AD (i.e. the main mix).</td></tr>
<tr><td colspan="2"><u>ImAc extension</u>
NO</td><td><u>Relevant standards</u>
MPEG DASH (ISO/IEC 23009-1:2014) [4]
DVB DASH Profile (ETSI TS 103 285 V1.1.1) [5]</td></tr>
</table>

<table>
<tr><td rowspan="2"># 2</td><td><u>Title</u>
Audio language</td><td><u>Location</u>
MPEG DASH manifest</td></tr>
<tr><td colspan="2"><u>Description</u>
Language of audio and (if applicable) AD</td></tr>
<tr><td colspan="3"><u>Usage in ImAc / Explanation</u>
The language is signalled using the @lang attribute on the AdaptationSet element.</td></tr>
</table>

ImAc extension	Relevant standards
NO	MPEG DASH (ISO/IEC 23009-1:2014) [4]

Sample

```
<AdaptationSet contentType="audio" lang="eng" …>

    ...

</AdaptationSet>
```

# 3	Title	Location
	AD mode	DASH manifest (ImAc extension)

Description

Differ between the following modes:

- classic
- static
- dynamic

Usage in ImAc / Explanation

ImAc defines an attribute for element ‘Representation’ in the MPEG DASH MPD. The attribute is added under an ImAc namespace.

Namespace declination: xmlns:imac="http://www.imac-project.eu"
@name: audioMode
Value enumeration: “classic”, “static”, “dynamic”

Note: Since the library used in the ImAc player does not allow to parse attributes with user defined namespaces, the namespace is dropped in the ImAc implementation.

ImAc extension	Relevant standards
YES	MPEG DASH (ISO/IEC 23009-1:2014) [4]

Sample

```
<Representation imac:audioMode="classic">
```

# 4	Title	Location
	AD volume	MPEG DASH manifest (ImAc extension)

Description

Volume level of AD (in relation to main audio)

Usage in ImAc / Explanation
ImAc defines an attribute for element ‘Representation’ in the MPEG DASH MPD. The attribute is added under an ImAc namespace. Namsepace declination: xmlns:imac="http://www.imac-project.eu" @name: audioDescriptionGain Value enumeration: “low”, “medium”, “high”

ImAc extension	Relevant standards
YES	MPEG DASH (ISO/IEC 23009-1:2014) [4]

Sample

```
<Representation imac:audioDescriptionGain="medium">
```

# 5	Title	Location
	Audio channel allocation	MPEG DASH (ImAc extension)

Description

Audio channel allocation and channel description

Usage in ImAc / Explanation

In a first ImAc implementation, the element AudioChannelConfiguration was used to specify the numbers of channels (we used either 2 for stereo or 4 for 1st order Ambisonics). When four channels were present, the ImAc player interprets the signal as Ambisonics audio. This is a proprietary solution, that is not compatible with any other player.

The signalization is realized by the AudioChannelConfiguration element, using the schemeIdUri “urn:mpeg:dash:23003:3:audio_channel_configuration:2011”.

Signalization for Ambisonics audio:
AudioChannelConfiguration @value: “4”

Signalization for Stereo audio:
AudioChannelConfiguration @value: “2”

It is intended to implement the proposed descriptor (refer to section 3.3.1) for Ambisonic signalization in the future.

ImAc extension	Relevant standards
YES	MPEG DASH (ISO/IEC 23009-1:2014) [4]

Sample

```
<AdaptationSet>
    //ambisonics

    <Representation>
        <AudioChannelConfiguration
            schemeIdUri="urn:mpeg:dash:23003:3:audio_channel_configuration:2011"
            value="4"/>
        ... // 4 channels audio
    </Representation>

</AdaptationSet>

<AdaptationSet>
    //stereo

    <Representation>
        <AudioChannelConfiguration
            schemeIdUri="urn:mpeg:dash:23003:3:audio_channel_configuration:2011"
            value="2"/>
        ... // 2 channels audio
    </Representation>

</AdaptationSet>
```

4. SPOKEN SUBTITLES

4.1. Motivation

Spoken subtitles (also called audio subtitles – AST) can be realized in two ways: either by using a text-to-speech engine at the client side, or by delivering the spoken subtitles as audio to the client.

The realization in the ImAc project is done by delivering audio due to some of the tested features that require audio rendering process that can better be done at broadcaster premises without facing the challenge for real time rendering. In this scenario, subtitles are delivered similarly to the Audio Description service. The main difference in the ImAc implementation is that the AST is mixed at the client side (receiver-mix) where all AD variations are pre-rendered before delivery (broadcast-mix).

In Deliverable D3.4 chapter 5.1 [1], we provided a gap analysis regarding the signalling of ImAc services and an overview of the required ImAc extensions. In this document, we describe possible solutions in detail, based on existing standards. Section 3.2 describes relevant data for audio and audio description in particular in the state-of-the-art MPEG DASH specification. The spoken subtitle service is similar but not identical to audio description. Standardized signalling data for audio description could be used for spoken subtitles as well (please refer to section 3.2 for state-of-the-art signalling for AD).

The requirements that arise from the ImAc services are:

1. Means to signal audio as spoken subtitle track in the delivery (streaming) format MPEG DASH.
2. Means to signal custom properties of audio streams in the delivery (streaming) format MPEG DASH.

A proposal for a possible signalling of AST is described in section 4.2. The signalling of audio properties (required to realize the ImAc audio modes) is equal to the signalling for audio description and described in section 3.3.2.

4.2. Proposed solutions for ImAc services

4.2.1. Signalling of spoken subtitles

As described in section 4.1, the player must be able to identify AST streams correctly in order to present the user a corresponding choice for the available services, in this case the AST. To ensure a correct rendering, the audio tracks that carry the AST must be identified as AST, and it must be signalled if this audio contains AST only (for receiver-mix) or AST mixed with the main audio (broadcast-mix). The latter can be signalled by DASH as described in section 3.2.1 using the Role element and the values “alternate” for broadcast mix AST and “commentary” for receiver mix AST.

Identifying audio as AST is not possible by the state-of-the-art DASH standard.

Proposal

Following the above standing requirement, we propose to signal AST audio on the DASH layer (or equivalent when using other streaming formats like HLS [14] or CMAF [15]).

For the usage of MPEG DASH, the Accessibility element should be used to carry this information. This element is intended to signal the type of access service which is our intention here.

We suggest to add an audio-subtitles identifier as a new value to the descriptor “urn:tva:metadata:cs:AudioPurposeCS:2007” for the Accessibility element. This descriptor uses integer values that are mapped to target audience groups. The value “1” is used to indicate the target group “for the visually impaired”, which is also used to signal audio description. The “visually impaired” users are also the target group of the AST service and an Accessibility element should be set accordingly. A second Accessibility element should carry the concrete type of the access service (AD or AST).

Sample implementation in ImAc:

The ImAc implementation defines a custom descriptor for the accessibility element, such that no conflicts with the existing standards occur. It is intended to keep this custom solution in place until a standardized solution is available. Signalization of audio subtitles in ImAc:

- Accessibility element with schemeIdUri “urn:imac:access-identifier:2019” and value “audio-subtitles”

4.3. Summary of signalling used in ImAc

The following tables provide an overview of the transportation of signalling and metadata for spoken subtitles as it is used in ImAc.

The meaning of most fields in the table are self-explaining, but few notes are listed below for a richer comprehension:

- The field “Location” describes from where the ImAc decoder (i.e. the ImAc player) retrieves the information. That means in what layer the information is located.
- The field “ImAc extension” indicates if the data is transported using standardized formats of if a format extension that was defined in the ImAc project is used.
- The numbering does not have any specific meaning and is just used for indexing the listed elements.

# 1	**Title** AST presence and target group	**Location** MPEG DASH manifest
Description Indication, if and where the spoken subtitle service is available in the stream.		
Usage in ImAc / Explanation The presence of AST within an audio track is signalled by the combination of two elements. Accessibility: Set to "audio-subtitles" using the schemeIdUri “urn:imac:access-identifier:2019” Role: The value "commentary" is suggested to use for Receiver mix AD and is used in ImAc to signal		

AST. AST is realized as receiver mix in ImAc.	
ImAc extension YES	**Relevant standards** MPEG DASH (ISO/IEC 23009-1:2014) [4] DVB DASH Profile (ETSI TS 103 285 V1.1.1) [5]

Sample

```
<AdaptationSet>

    <Role
        schemeIdUri="urn:mpeg:dash:role:2011"
        value="alternate"/>

   <Accessibility
        schemeIdUri="urn:imac:access-identifier:2019"
        value="audio-subtitles"/>
    …


</AdaptationSet>
```

# 2	**Title** Audio language	**Location** MPEG DASH manifest

Description

Language of the spoken subtitles

Usage in ImAc / Explanation

The language is signalled using the @lang attribute on the AdaptationSet element.

ImAc extension	**Relevant standards**
NO	MPEG DASH (ISO/IEC 23009-1:2014) [4]

Sample

```
<AdaptationSet contentType="audio" lang="eng" …>

    ...

</AdaptationSet>
```

<table>
<tr><td rowspan="2"># 3</td><td><u>Title</u>
AST mode</td><td><u>Location</u>
MPEG DASH manifest (ImAc extension)</td></tr>
<tr><td colspan="2"><u>Description</u>
Differs between the following modes:
• classic
• static
• dynamic</td></tr>
<tr><td colspan="3"><u>Usage in ImAc / Explanation</u>
Signalling of the AD mode is the same as described for Audio Description in section 3.4. ImAc defines an attribute for element ‘Representation’ in the MPEG DASH mpd. The attribute is added under an ImAc namespace.

Namespace declination: xmlns:imac="http://www.imac-project.eu"
@name: audioMode
Value enumeration: “classic”, “static”, “dynamic”

Note 1: Since the library used in the ImAc player does not allow to parse attributes with user defined namespaces, the namespace is dropped in the ImAc implementation.
Note 2: Just the classic and dynamic modes are being used and tested in ImAc for AST.</td></tr>
<tr><td colspan="2"><u>ImAc extension</u>
YES</td><td><u>Relevant standards</u>
MPEG DASH (ISO/IEC 23009-1:2014) [4]</td></tr>
<tr><td colspan="3"><u>Sample</u>
<code><Representation imac:audioMode="classic"></code></td></tr>
</table>

5. SIGN LANGUAGE

5.1. Motivation

The sign language video is presented in a similar manner than the subtitles in the ImAc project. That means that the sign language (SL) video is shown at a certain position on the current viewport and stays at that position fixed-to-screen.

As the subtitles, we have developed enhanced presentation options for the signer video, which are listed below:

- The signer video is enriched with cues that indicate the speaker location within the scene. This is done using the arrow indicator, an arrow, that is shown under the signer video, pointing in the direction of the related speaker in the scene, who is currently being interpreted. The presentation of a radar to indicate the position of the target speaker is also supported.
- Additionally, we implemented a presentation option where the speaker's name is written below the signer video to help the viewer with the speaker identification.
- Another presentation option allows to dynamically hide the signer video, when the sign language interpreter is currently inactive, i.e. not translating anybody, and to show it when the speaker is again active.

All these options require time related metadata transported along with the video.

Identified gaps in state-of-the-art standards

We found that current standards and specifications do not provide the required possibilities to realize all features of the developed presentation options. In this document, we describe possible solutions in detail, based on existing standards.

In particular, we have identified the following requirements that arise from the ImAc use cases[3]:

1. Means to signalize the sign language video intended size and position on screen
2. Means to signalize the 360° position of the speaker being interpreted
3. Means to signalize speaker identification information like speaker's name and/or a color code.
4. Means to signalize whether the sign language interpreter is currently active or inactive

Note: *Deliverable D3.4, chapter 5.1, provides a gap analysis regarding the signalling of ImAc services and an overview of the required ImAc extensions*.

Section 5.2 describes relevant data in the state-of-the-art MPEG DASH specification for the correct signalling of the signer video within DASH.

[3] The use cases that were tested with end users during various test phases are summarized in D5.2, second iteration [3]. Features for SL are listed in chapter 1.1.1.3.

Proposals for possible solutions to meet these requirements, based on DASH and TTML, are described in Section 5.3.

Section 5.4 provides a summary of relevant signalling data, as it is used in the ImAc implementation using existing standard means as well as ImAc specific extensions.

5.2. State of the art

5.2.1. Role element (draft status)

Location:
`AdaptationSet/ContentComponent` elements - as part of an MPD

Defined in:
- ISO/IEC FDIS 23009-1 [16], § 5.8.4.2, § 5.8.5.5
- DASH-IF Interoperability Points V4.2, [17] § 3.9.2

Description:

The Role element is an element in DASH to specify an arbitrary role for the affected component that is based on a specific scheme. The element can be used multiple times on a component.

`Role element – @schemeIdUri attribute`
This attribute specifies the scheme used for the role value. The DASH specification itself defines a simple role scheme (URI: `urn:mpeg:dash:role:2011`) which should be used here.

`Role element – @value attribute`

This attribute specifies the actual role value.

Regarding sign language, the following values from the mentioned scheme are used here:

- sign

5.2.2. Accessibilty element (draft status)

Location:
`AdaptationSet/ContentComponent` elements - as part of an MPD

Defined in:
- ISO/IEC FDIS 23009-1 [16], § 5.8.4.3, § 5.8.5.5
- DASH-IF Interoperability Points V4.2 [17], § 3.9.2, § 3.9.4.5

Description:

The Accessibility element is a generic element in DASH to specify an arbitrary accessibility property that is based on a specific scheme. The element can be used multiple times on a component.

Accessibility element – @schemeIdUri attribute

This attribute specifies the scheme used for the accessibility value. DASH itself doesn't define a scheme here. However the mentioned role scheme (URI: `urn:mpeg:dash:role:2011`) is used here, too.

Accessibility element – @value attribute

This attribute specifies the actual accessibility value.

- Regarding sign language, the following values from the mentioned scheme are used here:sign

5.3. Proposed solutions for ImAc services

The use case we have in ImAc requires time related metadata for the signer video. This is not supported sufficiently in current standards.

General solution using a timed sidecar metadata file

For a flexible and comprehensive carriage of time-related data, we propose the usage of a sidecar file carrying this data.

In the project, we are using a TTML based format with custom elements and attributes that were defined to describe and provide the required data. A TTML based format allows us to re-use the subtitle solution for indicating the position of the related speaker. Since TTML allows user defined extensions, all required information can be carried. Associating information with the media timeline can be realized by using TTML build-in timing mechanism.

As described in Section 5.1, the signer presentation options require time related metadata along with the signer video. Even if the signer video is transmitted as a continuous stream, it can be divided into logical segments, each segment containing a sign language interpreted snippet. The segmentation is only conceptional and will be realized via the TTML sidecar file, the video stream itself does not contain information regarding this segmentation.

We are transmitting metadata for each segment. That means that a new segment needs to be created each time the speaker changes, or if the sign language interpreter becomes active or inactive, or if the speaker move to another position in the 360º space. A segment is represented by a tt:p element in the TTML file. The following figure shows a sample file including two segments:

```
<?xml version="1.0" encoding="UTF-8"?>
<tt:tt xmlns:tt="http://www.w3.org/ns/ttml"
    xmlns:ttp="http://www.w3.org/ns/ttml#parameter"
    xmlns:tts="http://www.w3.org/ns/ttml#styling"
    xmlns:ttm="http://www.w3.org/ns/ttml#metadata"
    xmlns:imac="http://www.imac-project.eu"
    xmlns:ebuttm="urn:ebu:tt:metadata"
    xmlns:ebutts="urn:ebu:tt:style">
    <tt:head>
        <tt:metadata>
        </tt:metadata>
        <tt:styling>
            <tt:style xml:id="defaultStyle"/>
        </tt:styling>
        <tt:layout>
            <tt:region xml:id="default"
                tts:extent="80% 80%"
                tts:origin="10% 10%"/>
        </tt:layout>
    </tt:head>
    <tt:body>
        <tt:div style="defaultStyle">
            <tt:p begin="00:00:00.800"
                end="00:00:05.800"
                imac:equirectangularLongitude="30"
                region="default"
                xml:id="sign0">
                <tt:metadata>
                    <imac:speakerColorCode>
                        #FF0000
                    </imac:speakerColorCode>
                    <imac:speakerDisplayName>
                        Philip
                    </imac:speakerDisplayName>
                </tt:metadata>
            </tt:p>
            <tt:p begin="00:00:13.000"
                end="00:00:20.000"
                imac:equirectangularLongitude="-20"
                region="default"
                xml:id="sign1">
                <tt:metadata>
                    <imac:speakerColorCode>
                        #00FF00
                    </imac:speakerColorCode>
                    <imac:speakerDisplayName>
                        Dave
                    </imac:speakerDisplayName>
                </tt:metadata>
            </tt:p>
        </tt:div>
    </tt:body>
</tt:tt>
```

The TTML sidecar file will be signaled in DASH as a separate AdaptationSet. The separate data fields in the metadata file are described in the following sub-sections. At the DASH level, the file's AdaptationSet is described by the properties described in 5.3.1.

Current ImAc implementation

By the time of writing this document, the proposed solutions are not yet implemented in the ImAc tools. It is planned to move to the proposed solution in the future. For a detailed description of the current implementation, please refer to deliverable D3.5, Annex II [12]

5.3.1. AdaptationSet element of metadata file

Location

MPD/Period/AdaptationSet

Attribute @id

The @id attribute must be set on this AdaptationSet in order to identify it from the AdaptationSet that contains the signer video. The value can be arbitrary of type xs:string.

Attribute @contentType

The @contentType attribute defines the AdaptationSet content as "application". This is used since MPEG DASH specifies to use "application" in case the AdaptationSet content does not relate to any of the other media content types. (refer to RFC4288, $4.2 [13])

Attribute @mimeType

The @mimeType specifies the media content as TTML using the value "application/ttml+xml".

Role element - @schemeIdUri attribute

A Role element is added to the AdaptationSet, using a user defined descriptor which identifies the AdaptationSet purpose as signer metadata track. We use the ImAc descriptor used for other purposes in the project as well: urn:imac:access-identifier:2019

Role element - @value attribute

An additional value is defined particular for the identification of a signer metadata track: sign-metadata.

The following MPD sample shows the relevant information of the AdaptationSet related to the the sign language video and the metadata track.

Sample MDP snippet:

```
<AdaptationSet
    id="signerVideo"
    contentType="video"
    lang="sgn-gsg">

    <Role
        schemeIdUri="urn:mpeg:dash:role:2011"
        value="sign"/>

    <SupplementalProperty
        schemeIdUri="urn:imac:signer-metadata-adaptation-set-id:2019"
        value="signerMetadata" />

    <Representation
        id="signerHQ">

        <BaseURL>
            ...
        </BaseURL>

    </Representation>

    <Representation
        id="signerLQ">

        ...

    </Representation>

</AdaptationSet>

<AdaptationSet
    id="signerMetadata"
    contentType="application"
    mimeType="application/ttml+xml">

    <BaseURL>
        ...
    </BaseURL>

    <Role
        schemeIdUri="urn:imac:access-identifier:2019"
        value="sign-metadata" />

</AdaptationSet>
```

5.3.2. Signer video size and position

At this stage, no suggestions are made by the ImAc project regarding handling the size and position of the signer video within the 360° scene.

ImAc implementation:

Video size and horizontal position is set by user via personalization features of the player UI. The vertical position is fixed by the player implementation. Just in case that subtitles and sign language are enabled at the same time (supported feature by the player), if the subtitles position are set to top, then the sign language window will be also moved to top, as subtitling is considered the master service in such a case.

5.3.3. Location of the speaker

For each segment of the sign language interpreter, the horizontal direction of the related speaker needs to be identified. The information is carried in the TTML sidecar file as described in section 5.3.

Proposed solution

We propose to re-use the user defined attributes introduced in section 2.4.2. Please refer to this section for detailed information.

5.3.4. Speaker identification information

For each segment of the sign language interpreter, information on the related speaker needs to be described. The information is carried using the TTML sidecar file as described in section 5.3.

Proposed solution

The data used for speaker identification in the ImAc implementation contains:

- A color code
- The speaker's name

This solution is determined by the results from the pre-pilot tests 2, in which this approach was compared to the use of emojis / pictograms to represent speakers' faces, and was shown to be preferred by users.

We are using the tt:metadata element to carry the above listed information within a tt:p element. Custom metadata elements have been defined under a user defined namespace:

- Namespace used: http://www.imac-project.eu
- Abbreviation used: imac

The color code is described as a RGB color triple in hex notation in the element "imac:speakerColorCode"

The speaker's name is put as plain text content in the element "imac:speakerDisplayName".

Sample TTML snippet (ImAc implementation):

```
<tt:p ... >
    <tt:metadata>
        <imac:speakerColorCode>
            #FF0000
        </imac:speakerColorCode>
        <imac:speakerDisplayName>
            Philip
        </imac:speakerDisplayName>
    </tt:metadata>
</tt:p>
```

5.3.5. Status of sign language interpreter (active / inactive)

The sign language interpreter status describes when a sign language interpreter is active. When the interpreter is inactive, the signer video may show him/her in a neutral position, or the video may show any other placeholder between the last and the next sign language segment, or directly the video window can be hidden (which can contribute to a higher immersion). If the produce sign language stream is discontinuous, then a continuous stream can be delivered by inserting black frames for the inactive periods.

The information of the signer activity is carried using the TTML sidecar file as described in section 5.3.

Proposed solution

The status is indicated by the presence of a tt:p element within the TTML sidecar file. If a tt:p element is active at a given time, the sign language interpreter is active, if no tt:p element is active at that time, the SL interpreter is inactive.

5.4. Summary of signalling used in ImAc

The following tables provide an overview of the transportation of signalling and metadata for the signer video, as used in ImAc.

The meaning of most fields in the table are self-explaining, but few notes are listed below for a richer comprehension:

- The field “Location” describes from where the ImAc decoder (i.e. the ImAc player) retrieves the information. That means in what layer the information is located.
- The field “ImAc extension” indicates if the data is transported using standardized formats of if a format extension defined in the ImAc project is used.
- The numbering does not have any specific meaning and is just used for indexing the listed elements.

# 1	**Title** Signer presence	**Location** MPEG DASH manifest
Description Indicate if/where signer video is available in the stream		
Usage in ImAc / Explanation The presence of a SL video track is signalized by a video AdaptationSet element that contains the "Role" element with the following properties: Role: Value set to “sign” using the schemeIdUrn “urn:mpeg:dash:role:2011” Accessibility: Value set to “sign” using the schemeIdUrn “urn:mpeg:dash:role:2011” The value “sign” is new in the currently drafted update of MPEG-DASH (urn:mpeg:dash:role:2011 combining 3rd edition and Amd5 supports value "sign")		
ImAc extension NO	**Relevant standards** MPEG DASH (ISO/IEC 23009-1:2014) [4]	

Sample

```
<AdaptationSet>

    <Role
        schemeIdUri="urn:mpeg:dash:role:2011"
        value="sign"/>

    <Accessibility
        schemeIdUri=" urn:mpeg:dash:role:2011"
        value="sign"/>

    <Representation>

        ...

    </Representation>

</AdaptationSet>
```

# 2	**Title** Language	**Location** MPEG DASH manifest
Description Signalization of the language in which the SL interpreting is provided		
Usage in ImAc / Explanation		

The language is signalized by the @lang attribute of the AdaptationSet element. ImAc uses the suggested prefix “sgn” followed by a dash and the three-digit sign language abbreviation to indicate the language of a sign language video track.

ImAc extension	**Relevant standards**
NO	MPEG DASH (ISO/IEC 23009-1:2014) [4]

Sample

```
<AdaptationSet
     lang="sgn-gsg" … >
     ...
</AdaptationSet>
```

# 3	**Title**	**Location**
	Position and size	Player implementation

Description

Parameters of video element in the 3D scene: Video size and vertical and horizontal position of the signer video within the field of view

Usage in ImAc / Explanation

Video size and horizontal position is set by user preferences. The vertical position is fixed by the player implementation, or determined by the subtitles position if both services are simultaneously active.

ImAc extension	**Relevant standards**
-	-

# 4	**Title**	**Location**
	Speaker identification	TTML metadata file

Description

Name and color of the speaker that is currently being interpreted.

Usage in ImAc / Explanation

The tt:metadata element carries speaker’s name and color code within a tt:p element. Custom metadata elements have been defined under a user defined namespace:

- Namespace used: http://www.imac-project.eu
- Abbreviation used: imac

The color code is described as a RGB color triple in hex notation in the element “imac:speakerColorCode”

The speaker's name is put as plain text content in the element "imac:speakerDisplayName".

ImAc extension	**Relevant standards**
YES	TTML1 (Timed Text Markup Language) [8]

Sample

```
<tt:p ... >
    <tt:metadata>
        <imac:speakerColorCode>
            #FF0000
        </imac:speakerColorCode>
        <imac:speakerDisplayName>
            Philip
        </imac:speakerDisplayName>
    </tt:metadata>
</tt:p>
```

# 5	**Title**	**Location**
	360° speaker position	TTML metadata file

Description

Signalization of the speaker horizontal direction.

Usage in ImAc / Explanation

The attribute in a user defined namespace is set on tt:p elements.

Longitude value rage: [-180; 180]

Namespace declination: xmlns:imac="http://www.imac-project.eu"

The center point [0; 0] is defined as the point on the sphere where the center of the equirectangular video texture is projected. Viewing directions up and left result in increasing latitude and longitude values.

ImAc extension	**Relevant standards**
YES	TTML1 (Timed Text Markup Language) [8]

Sample

```
<tt:p
    xml:id="s1"
    region="bottom"
    begin="00:00:01.000"
    end="00:00:04.000"
    imac:equirectangularLongitude="15">

    ...

</tt:p>
```

# 6	**Title** Status of SL interpreter	**Location** TTML metadata file
Description Signalize if the sign language interpreter is currently active or inactive.		
Usage in ImAc / Explanation The status is indicated by the presence of a tt:p element within the TTML sidecar file. If a tt:p element is active at a given time / period, then the sign language interpreter is active, if no tt:p element is active at that time, then the SL interpreter is inactive.		
ImAc extension YES	**Relevant standards** TTML1 (Timed Text Markup Language) [8]	

6. REFERENCES

[1] ImAc deliverable D3.4 - Accessibility Interface, Revision 1.0, August 2019, http://www.imac-project.eu/documentation/deliverables/

[2] ImAc deliverable D3.6 – Integration and Testing Report, Revision 2.0, September 2019, http://www.imac-project.eu/documentation/deliverables/

[3] ImAc deliverable D5.2 v0.6 - Pilot evaluation methodology plan, June 2019, http://www.imac-project.eu/documentation/deliverables/

[4] ISO/IEC 23009-1:2014 Information technology – Adaptive dynamic streaming over HTTP (DASH), Part 1: Media presentation description and segment formats; second edition May 2014, ISO/IEC 23009-1:2014

[5] ETSI TS 103 285 V1.1.1 - Digital Video Broadcasting (DVB); MPEG-DASH Profile for Transport of ISO BMFF Based DVB Services over IP Based Networks, May 2015, https://www.etsi.org/deliver/etsi_ts/103200_103299/103285/01.01.01_60/ts_103285v010101p.pdf

[6] ETSI TS 102 822-3-1 V1.9.2 - Broadcast and On-line Services: Search, select, and rightful use of content on personal storage systems ("TV-Anytime"); Part 3: Metadata; Sub-part 1: Phase 1 - Metadata schemas, March 2016, https://www.etsi.org/deliver/etsi_ts/102800_102899/1028220301/01.09.02_60/ts_10282220301v010902p.pdf

[7] W3C IMSC 1.0.1 - TTML Profiles for Internet Media Subtitles and Captions 1.0.1 (IMSC1), W3C Recommendation, April 2018, https://www.w3.org/TR/ttml-imsc1.0.1/

[8] W3C TTML1 - Timed Text Markup Language 1 (TTML1) (Third Edition), November 2018, https://www.w3.org/TR/ttml1/

[9] ITU-R BT.2207-3, Accessibility to broadcasting services for persons with disabilities, October 2017, https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-BT.2207-3-2017-PDF-E.pdf

[10] EBU-TT Classification Scheme , EBU MIM-XMLSubs group, September 2015, https://www.ebu.ch/metadata/cs/EBU-TTContentTypeCS.xml

[11] ImAc deliverable D5.4 v0.4 - Pilot evaluation report, November 2018, http://www.imac-project.eu/documentation/deliverables/

[12] ImAc deliverable D3.5 – Player, Revision 2.1, July 2019, http://www.imac-project.eu/documentation/deliverables/

[13] RFC4288, Media Type Specifications and Registration Procedures, December 2005 https://tools.ietf.org/html/rfc4288

[14] RFC8216 - HTTP Live Streaming, August 2017, https://tools.ietf.org/html/rfc8216

[15] ISO/IEC CD 23000-19 Common Media Application Format, June 2016, https://mpeg.chiariglione.org/sites/default/files/files/standards/parts/docs/w16186.zip

[16] ISO/IEC FDIS 23009-1, Information technology - Dynamic adaptive streaming over HTTP (DASH) -- Part 1: Media presentation description and segment formats, August 2019, https://www.iso.org/standard/75485.html

[17] DASH-IF Interoperability Points V4.2, April 2018, https://dashif.org/docs/DASH-IF-IOP-v4.2-clean.htm

[18]RFC5646 – Tags for Identifying Languages, September 2009, https://tools.ietf.org/html/rfc5646

<END OF DOCUMENT>