



# A11y\*: not on TV

*AKA: Why Hacking is important in Research*

*We will focus on subtitling, but applies to all access services.*

LeadMe Winter School 1  
Chris Hughes  
c.j.hughes@Salford.ac.uk

\*A11Y stands for **accessibility** – the 11 refers to the eleven letters between the A and the Y in the word "**accessibility**".

# In this workshop

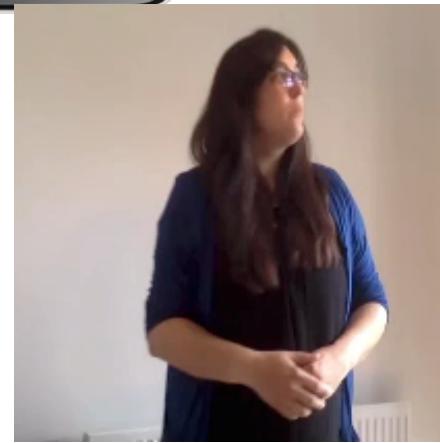
A ~~hackers~~ *engineers* story to contributing to immersive subtitling:

1. Away from the TV – Screens are all around us now
2. Stepping into the video
3. First steps are important but don't go far enough
4. Rapid Prototyping demos (hacking)

Disclaimer: I am a Technologist who has been involved in user tests, not a user test expert!

# Screens Are all around

(A Case study in responsive subtitling)



00:01.000

00:01.000

00:01.410

00:01.810

Here they are!

00:01.410

00:01.810

00:02.220

00:02.220

00:01.000

00:01.000

Here

00:01.410

they

00:01.810

are!

00:01.410

00:01.810

00:02.220

00:02.220

Here they are!

Who the hell is warbling  
like that outside our door?

Here  
they  
are!

Who  
the  
hell  
is  
warbling  
like  
that  
outside  
our  
door?

00:01.000

00:02.220

Here they are!

00:06.510

00:07.720

Who the hell is

00:07.720

00:08.940

warbling like that

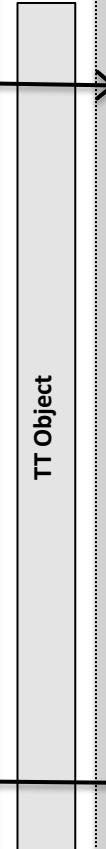
00:08.940

00:10.160

outside our door?

### IMSC.js

```
<tt:p begin="00:00:01.000" end="00:00:02.220" xml:id="C1">
  <tt:span style="S1">Here they are!</tt:span>
</tt:p>
<tt:p begin="00:00:06.100" end="00:00:10.160" xml:id="C2">
  <tt:span style="S2">Who the hell is warbling</tt:span>
  <tt:br/>
  <tt:span style="S2">like that outside our door?</tt:span>
</tt:p>
```



### responsiveSubs.js

```
[00:01.000 -> 00:01.410] (S1) "Here"
[00:01.410 -> 00:01.810] (S1) "they"
[00:01.810 -> 00:02.220] (S1) "are!"
[00:06.100 -> 00:06.510] (S2) "Who"
[00:06.510 -> 00:06.910] (S2) "the"
[00:06.910 -> 00:07.320] (S2) "hell"
[00:07.320 -> 00:07.720] (S2) "Is"
[00:07.720 -> 00:08.130] (S2) "warbling"
[00:08.130 -> 00:08.540] (S2) "like"
[00:08.540 -> 00:08.940] (S2) "that"
```

```
<tt:p begin="00:00:01.000" end="00:00:02.220" xml:id="C1">
  <tt:span style="S1">Here they are!</tt:span>
</tt:p>
<tt:p begin="00:00:06.100" end="00:00:08.940" xml:id="C2">
  <tt:span style="S2">Who the hell is</tt:span>
  <tt:br/>
  <tt:span style="S2">warbling like that</tt:span>
</tt:p>
<tt:p begin="00:00:08.940" end="00:00:10.160" xml:id="C2">
  <tt:span style="S2">outside our door?</tt:span>
</tt:p>
```

```
[00:01.000 -> 00:02.220] (S1) "Here they are!"
[00:06.100 -> 00:10.160] (S2) "Who the hell is warbling like
  that outside our door?"
```

```
[00:01.000 -> 00:02.220] (S1) "Here they are!"
[00:06.100 -> 00:07.720] (S2) "Who the hell"
[00:07.724 -> 00:08.940] (S2) "warbling like that"
[00:08.942 -> 00:10.160] (S2) "outside our door?"
```



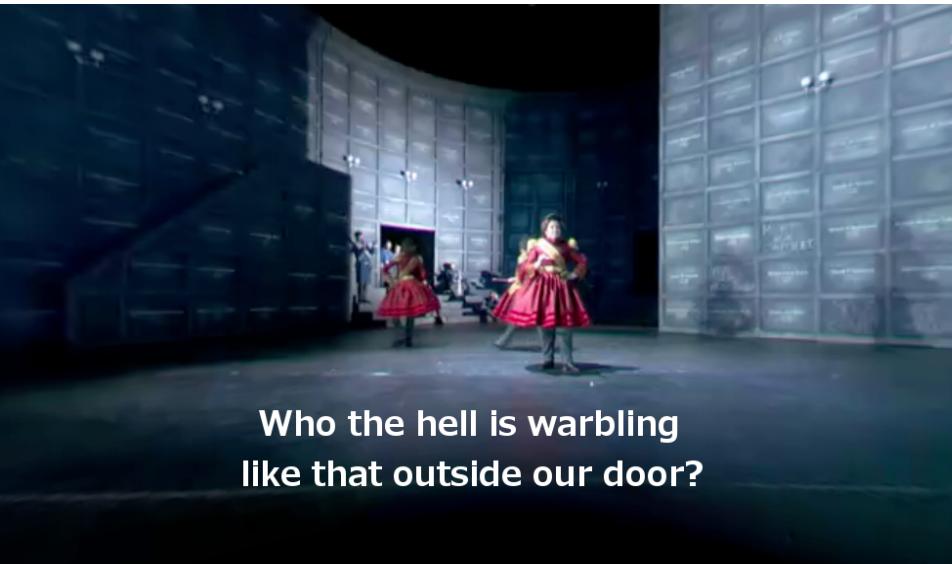
• 28

00:00:24.36

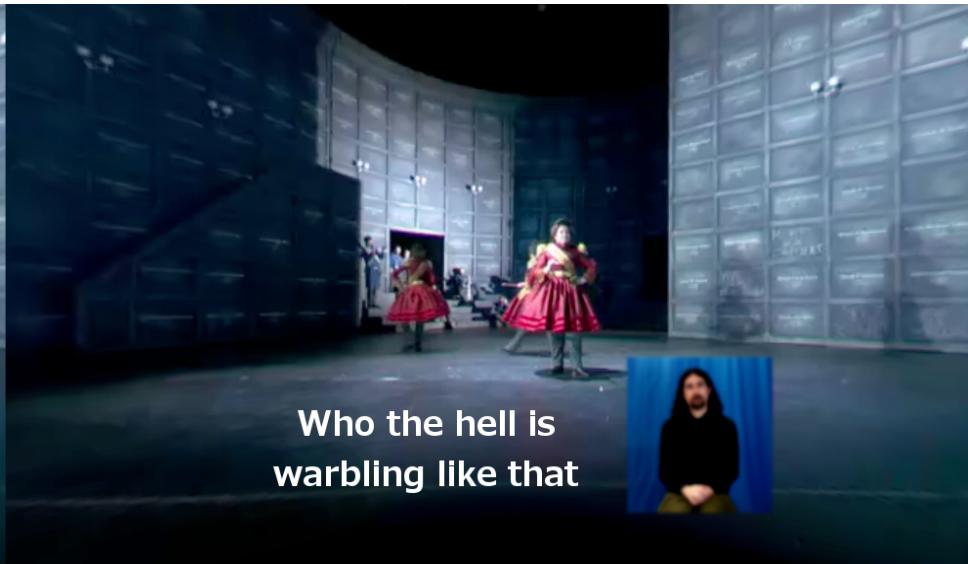


**The very same!  
He's very rash!**

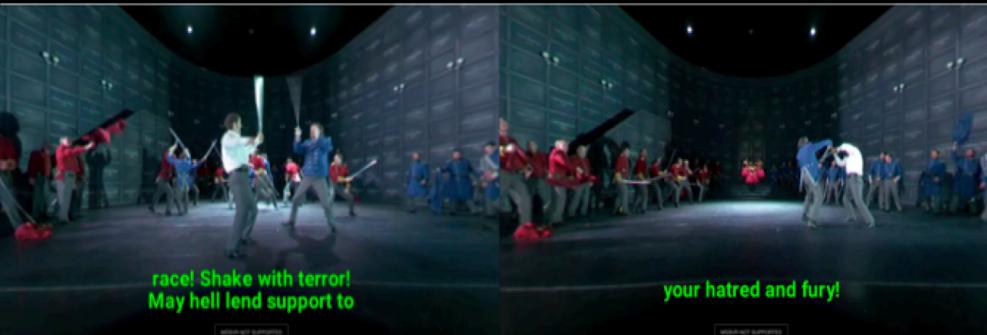
WEBVR NOT SUPPORTED



**Who the hell is warbling  
like that outside our door?**



**Who the hell is  
warbling like that**



# Try it...

<http://www.chxr.org/responsive>



# References

- Armstrong, Mike, Brown, Andy, Crabb, Michael, Hughes, CJ , Jones, Rianne and Sandford, James 2017, 'Understanding the diverse needs of subtitle users in a rapidly evolving media landscape' , SMPTE Motion Imaging Journal, 125 (9) , pp. 33-41.
- Crabb, M, Jones, R, Armstrong, M and Hughes, CJ 2015, Online news videos : the UX of subtitle position , in: 17th International ACM SIGACCESS Conference on Computers & Accessibility, 26-28 October 2015, Lisbon, Portugal.
- Hughes, CJ , Armstrong, M, Jones, R and Crabb, M 2015, Responsive design for personalised subtitles , in: The 12th Web for All Conference, 18-20 May 2015, Florence, Italy.
- Hughes, CJ and Armstrong, M 2015, Automatic retrieval of closed captions for web clips from broadcast TV content , in: 2015 NAB Broadcast Engineering Conference, 6-11 April 2015, Las Vegas, USA.

# Stepping into the video

Immersive Accessibility (ImAc)  
Project

# Immersive content

**What type of immersive content are workshop participants familiar with?**

# **Immersive content**

**Virtual Reality (including 360º videos)**

**Augmented Reality**

**Mixed Reality**

# Challenges: home users

**How to guide users to the person speaking?**

**Where to position the subtitles?**

**What is the comfort viewing field?**

**How to deal with non-speech information?**

**What features to include in the user interface?**

# **Challenges: professionals**

**How to show the 360º scene on the editor?**

**How to position subtitles?**

**How to visualise and test the result?**

# User-centric approach

1. **WHO?** Define your users.
2. **WHAT?** Define what you want to develop: editing tools, player, access services. Different scenarios.
3. **HOW?** Define the methodology to gather data from users: qualitative versus quantitative methods.
4. Implement and test in different iterations.
5. Keep track of everything (requirement list).

# ImAc Project

# Introduction to ImAc

**ImAc** (Immersive Accessibility): EU H2020 project that is exploring how accessibility services can be efficiently integrated with immersive media

- **Access Services:** subtitles, audio description, sign language
- **Immersive Media:** omnidirectional video (i.e. 360º) and audio

## Premises:

- Accessibility is a must for **e-inclusion**
- Accessibility must **not** be considered as an **afterthought**, but as a key aspect in the specification and deployment of services
- Keep **compatibility** with current standard technologies / formats
- **User-Centric Methodology**



# Introduction to ImAc

## Communication

- Website: <http://www.imac-project.eu/>
- Twitter: @ImAcProject
- Cross-disciplinary team

Partners



Funded by



# ImAc Objectives

**OBJ1.** Create accessible and fully personalized services for all citizens

**OBJ2.** Deliver novel resources for the broadcasting industry to provide adapted content ensuring accessibility in immersive environments

**OBJ3.** Demonstrate the tools and platforms in open pilots

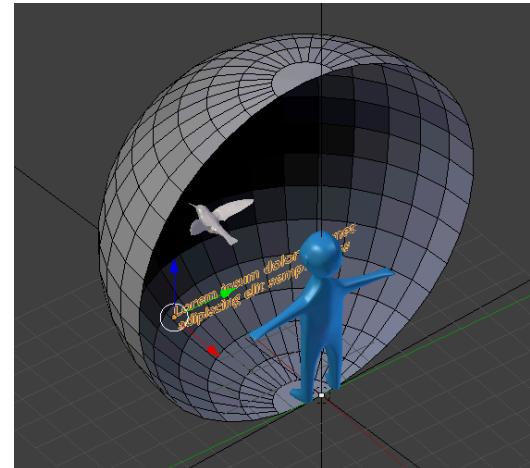
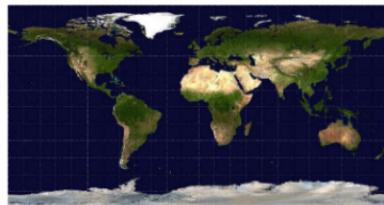
**OBJ4.** Work towards standardization of accessibility data in an immersive content environment

**OBJ5.** Maximize impact on society delivering real and useful solutions

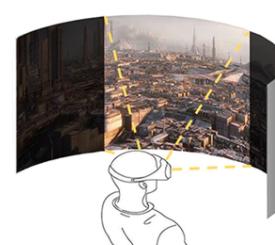


# 360° Video

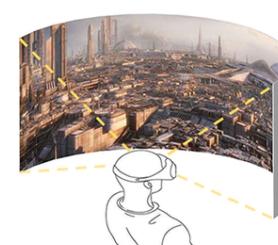
Time... and space!



100°



210°

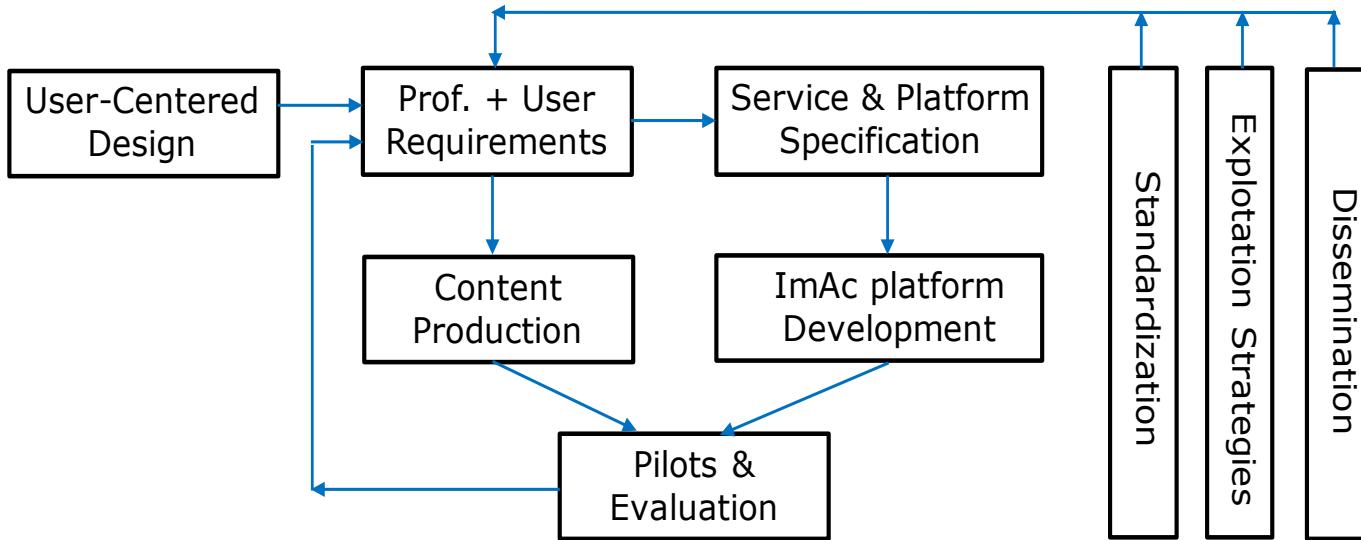


# ImAc Methodology

## Three Key Pillars:

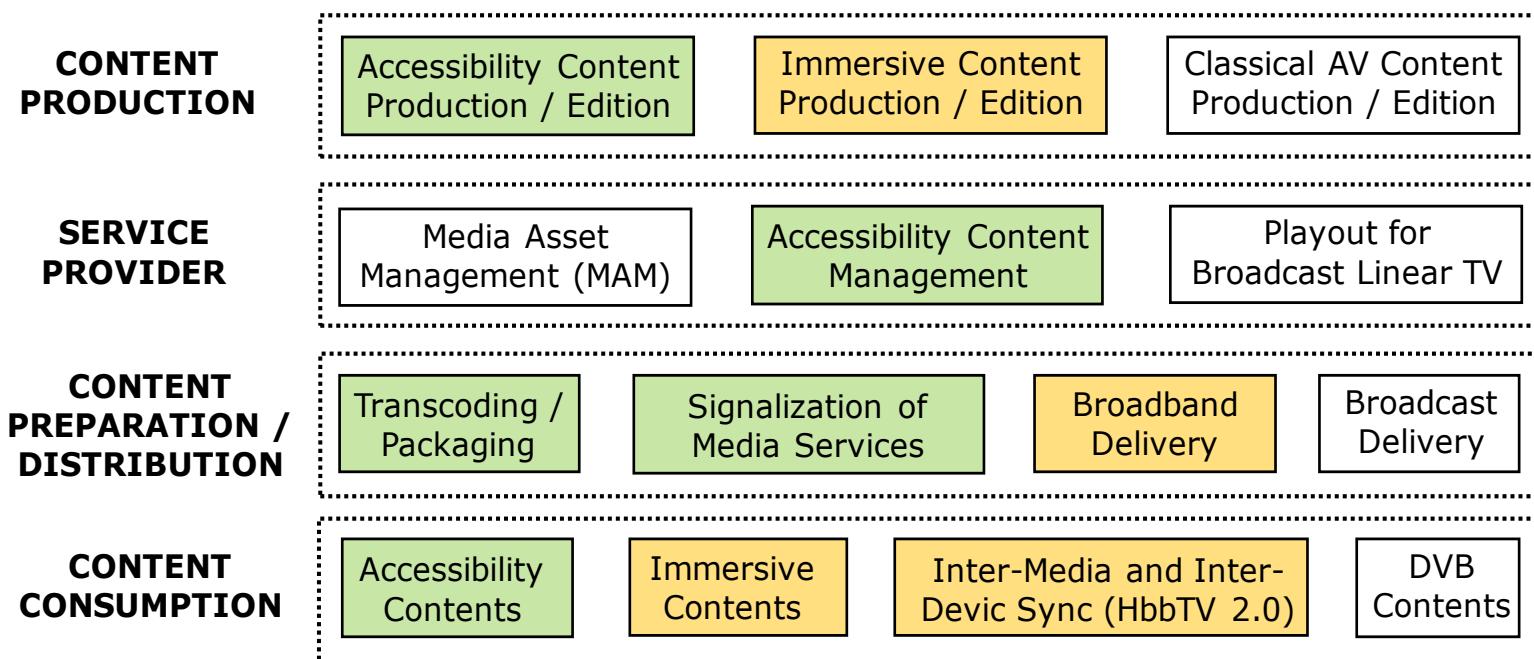
- 1) Requirement Gathering
- 2) Development and Integration
- 3) Validation and dissemination

*“Design for users, with users”*

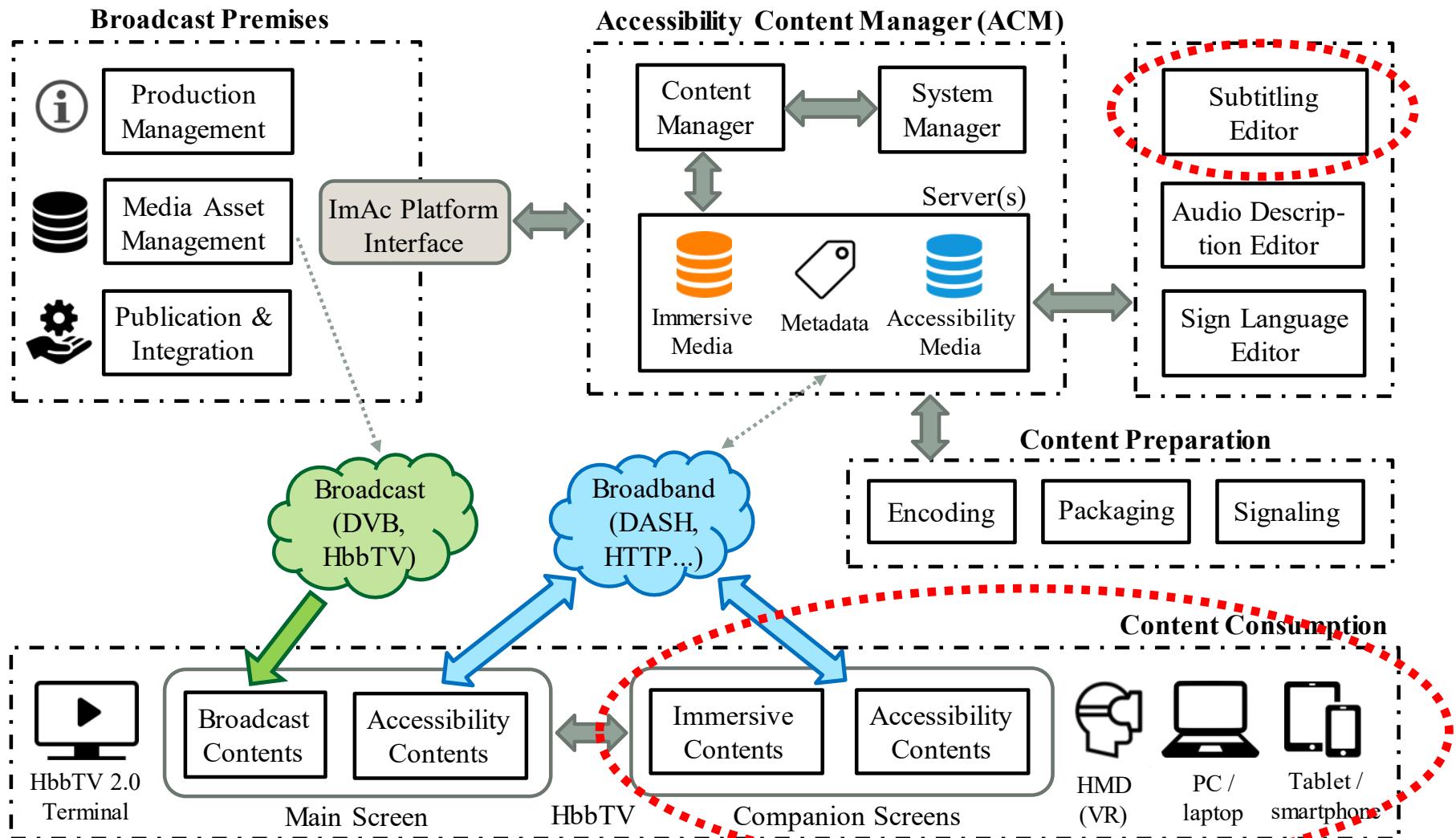


# ImAc Platform

## Main Parts:



# ImAc Platform



# ImAc Outputs

## **Key components of the ImAc platform**

- Accessibility Content Manager
- Edition tools (SaaS)
- Open-Source VR360 player

## **Dissemination:**

- High-Impact Publications
- Events / Fairs / Workshops

## **Standardization:**

- Contributions to MPEG, W3C, ITU, ISO...

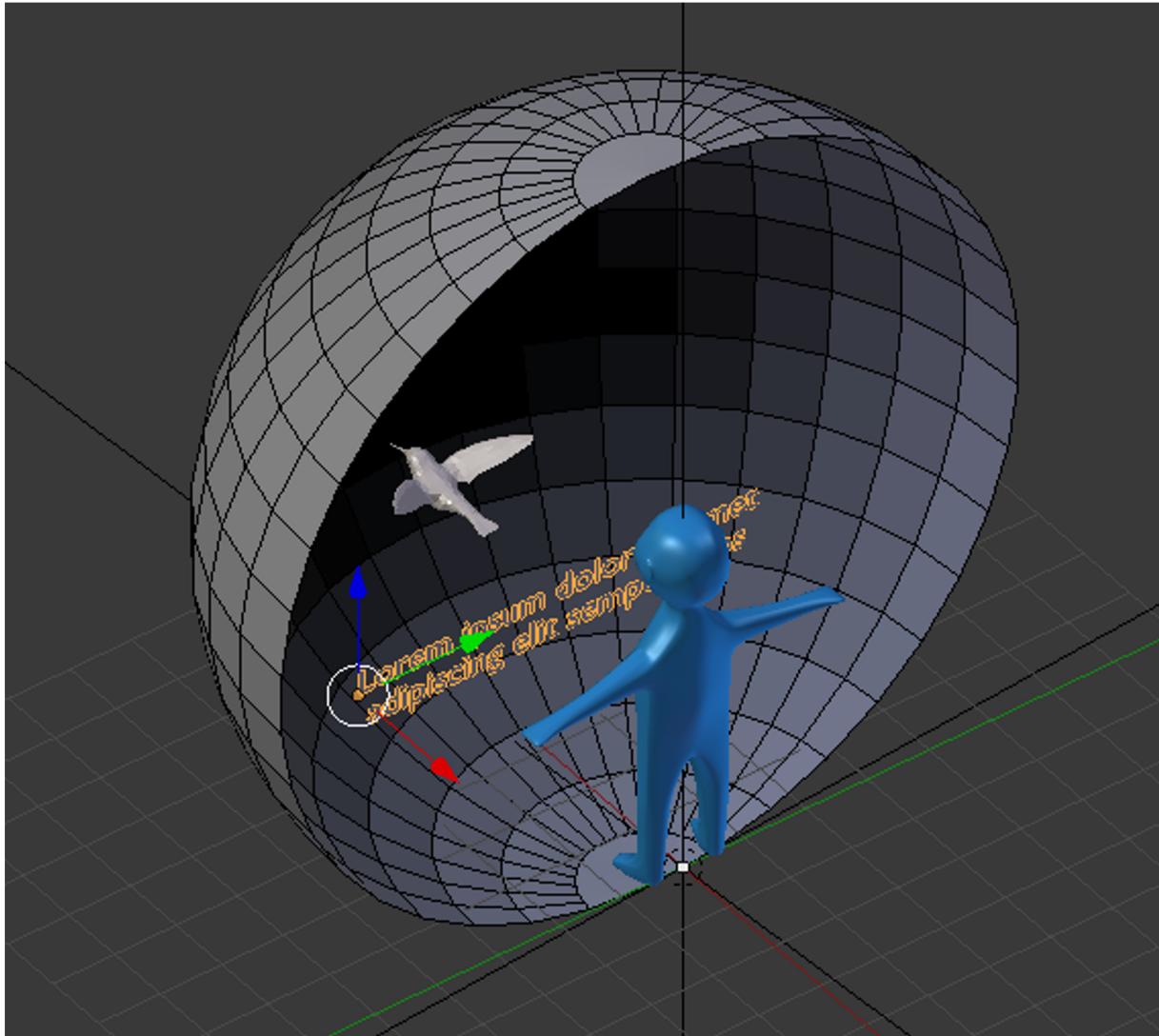
## **Benefits:**

- Contributions and insights very welcome by interested agents (research community, end-users, professionals, stakeholders.)



# ImAc Defining Requirements

# Introduction

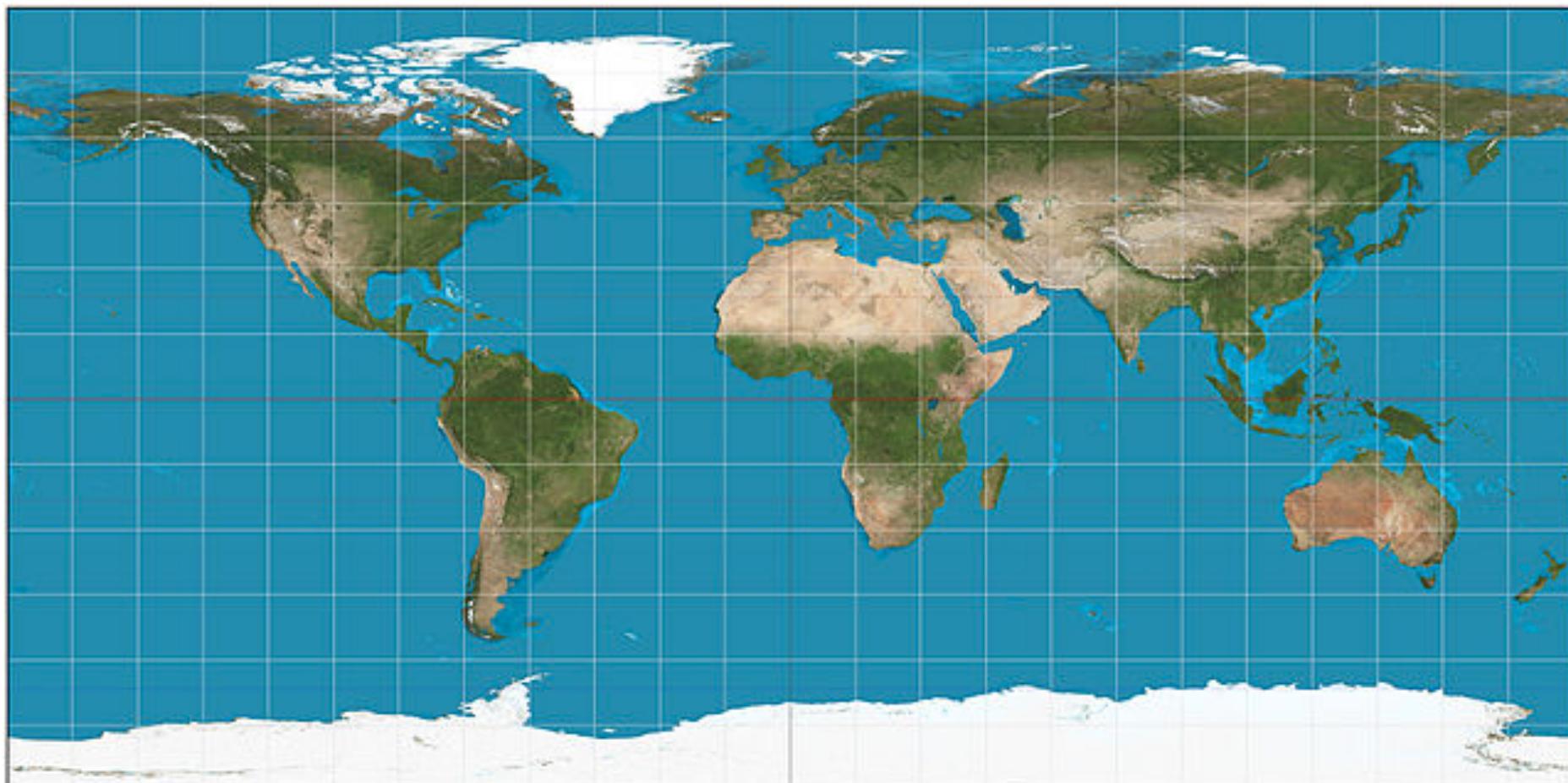


# Overview

- **Defining Requirements for Immersive Subtitling**
  - Challenges
  - Subtitling Workflow
  - Requirements for Editor, ACM, Distribution, Player
  - Presentation
- **Architecture and workflow**
  - Architecture
  - Delivery
  - Editor
  - Player
- **Conclusions**



# 360° scenery



Source: Wikipedia, [https://en.wikipedia.org/wiki/Equirectangular\\_projection](https://en.wikipedia.org/wiki/Equirectangular_projection)



# Subtitling

## Main challenges

- Comfort & Readability
  - Especially for VR glasses
  - Where can subtitles be rendered on the screen (safe area)?
  - What fonts and text sizes are reasonable?
- Speaker identification
  - How does the viewer know who is speaking?
  - How can the user keep orientation in the scene?

# Comfort & Readability



Image quality falls off towards the edges

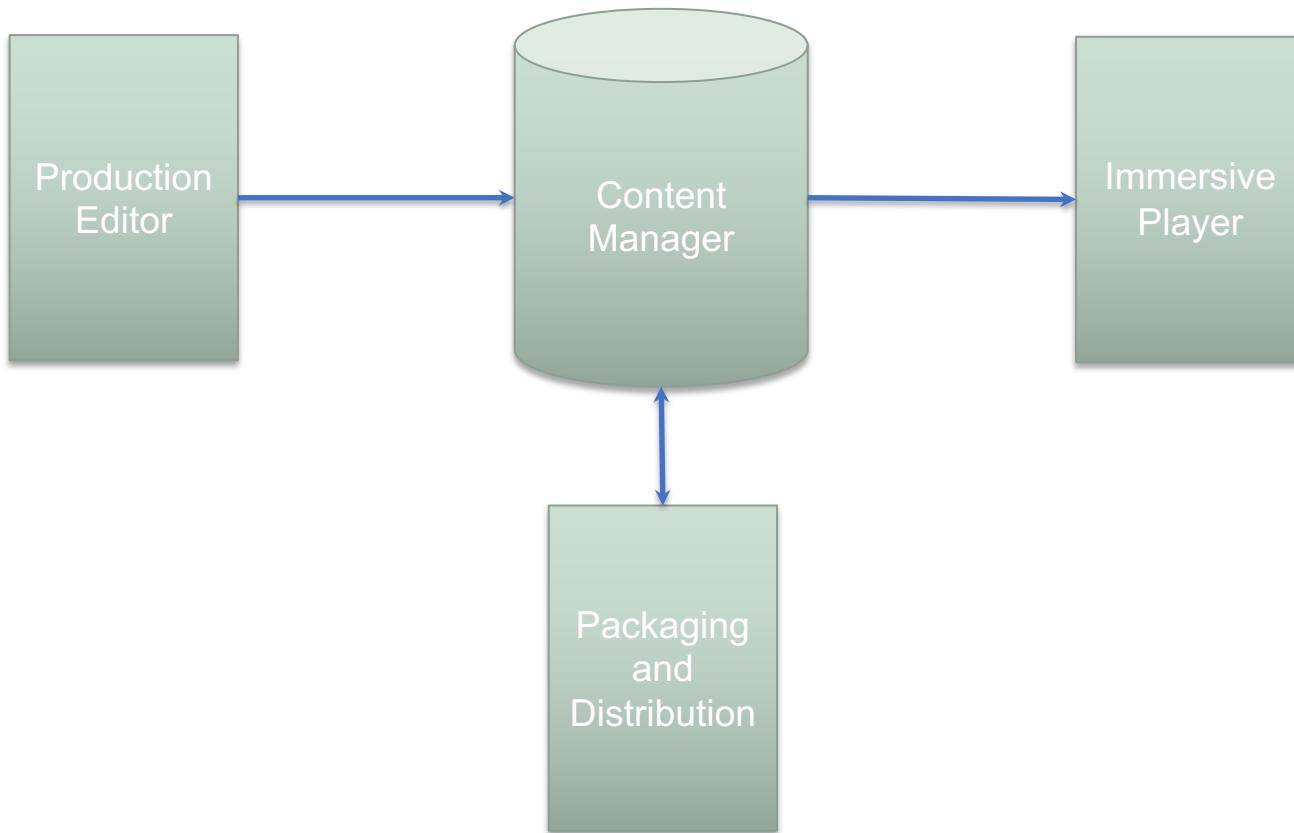
(Only exemplary, does not represent real image quality)

Photo through lens of the Oculus Go

# Speaker Identification



# Subtitling Workflow



# Requirements - Editor

- Allow the professional user to edit subtitles in both 'vr-mode' and 'normal mode'.
- Video should be managed via the ACM.
- Subtitling files should also be handled by the ACM, but it is foreseen to allow directly loading and saving subtitling content files from the computer hard drive on the second iteration.
- A 360 preview player for the low-resolution video playback and monitoring will be embedded in the editor with the following The video will be able to be viewed either as Equirectangular or VR view.



# Requirements - ACM

- Assign metadata to the asset with the accessibility contents, such as the ID needed to match the asset with external asset management systems.
- Prepare the final files for the packager using the asset's contents, such as a new mixed audio file created from the audio description file and the video file.
- Trigger the packaging by validating all the files required.
- The accessibility content manager web service will be used by the content packaging and distribution module to initiate and notify its processes.



# Requirements - Distribution

- Encode the content (compress the audio and video).
- Package the content inside a suitable container for delivery, and it will ensure that the content is segmented appropriately.
- This module will ensure a proper delivery that includes the signalling of the metadata according the editors' wishes.
- Push the content to the appropriate network (whether using streaming protocols or file-based caching infrastructures).

# Requirements - Player

- The UI needs to take into account the sensorial capacities of the target users.
- According to the focus group tests, different presentation modes and personalization options were identified.
- Colours to distinguish between different speakers:
  - The same colours will be used as per broadcaster's guidelines (3-4 colours).
  - The use of specific colour palettes for colour-blind people will also be considered.
- Maximum amount of text for one subtitle:
  - The TV broadcast subtitle services are still based on the teletext limitations as well as on viewer behaviour, including their reading speed. That implies that the number of characters is limited to about 37 characters per line and it is suggested to use not more than 2 lines for a subtitle.

# Requirements - Player

- Without VR glasses, we assume the same ratio (16:9) and the safe area for interactive elements defined by the EBU guideline EBU R95 (<https://tech.ebu.ch/docs/r/r095.pdf>) in a HD format.
- The head mounted displays have different maximum visual fields defined as FoV (Field of View). The comfortable area for viewing is still being tested.



# Presentation Modes

- Adding arrows
- Using sided
- Adding a compass or wind



# Presentation Modes

- Notifications for Sound

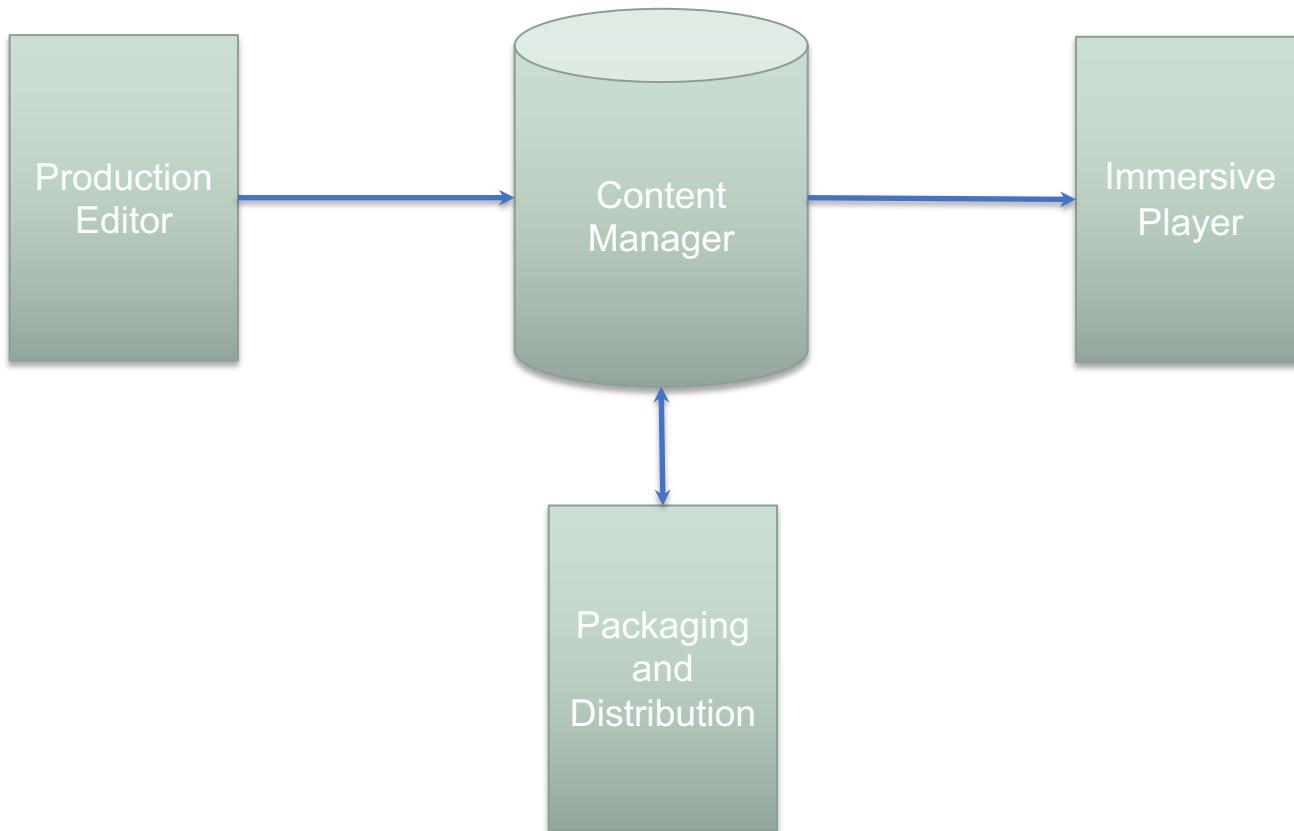


# Personalisation

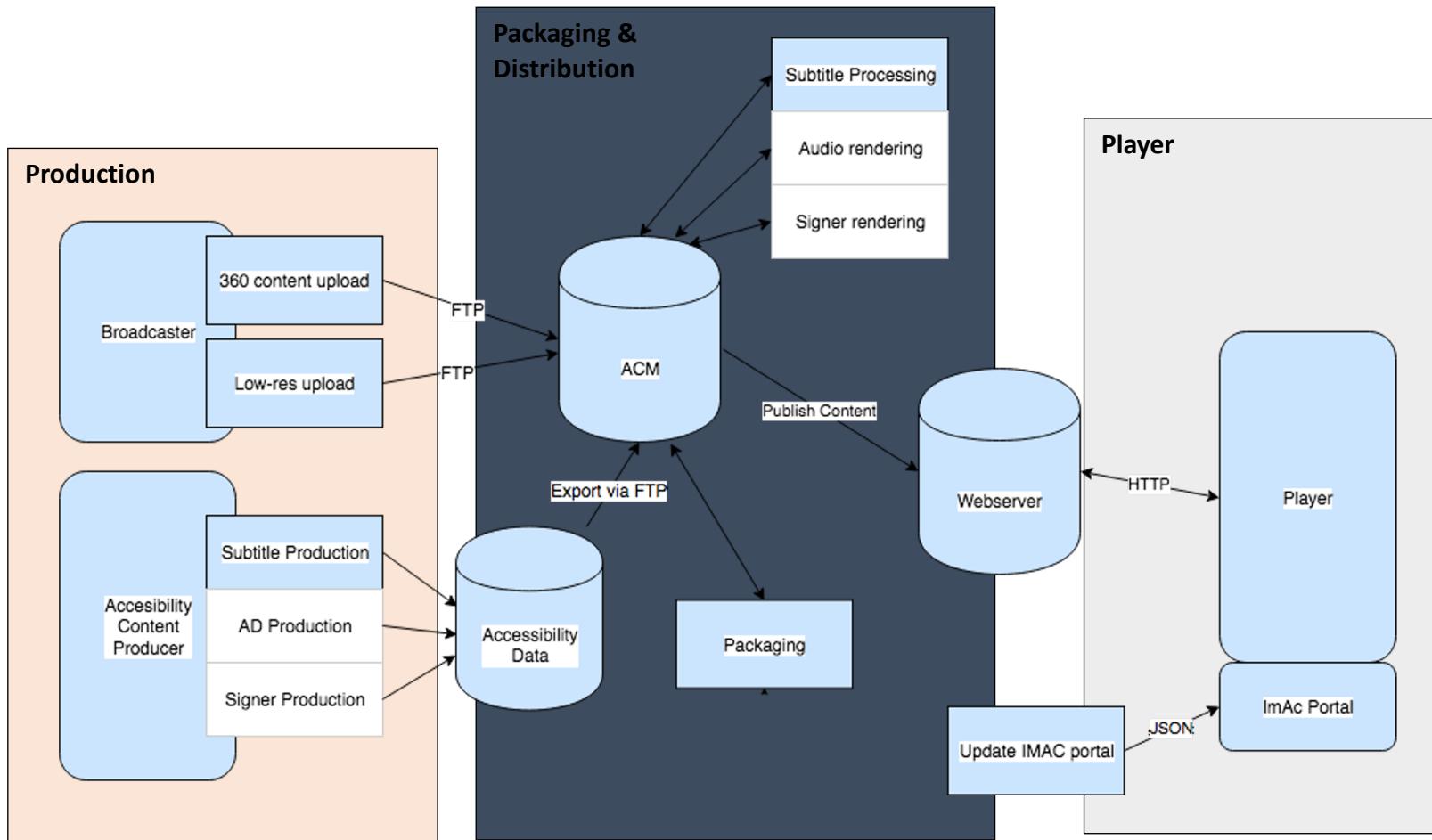
- Different font sizes (e.g. small, medium and large)
- Select between different backgrounds:
  - 1) semi-transparent box (80% opacity)
  - 2) outline (2px for each font size)
- Adding the subtitles below a scaled down video area.

ImAc Architecture

# Implementation



# Architecture

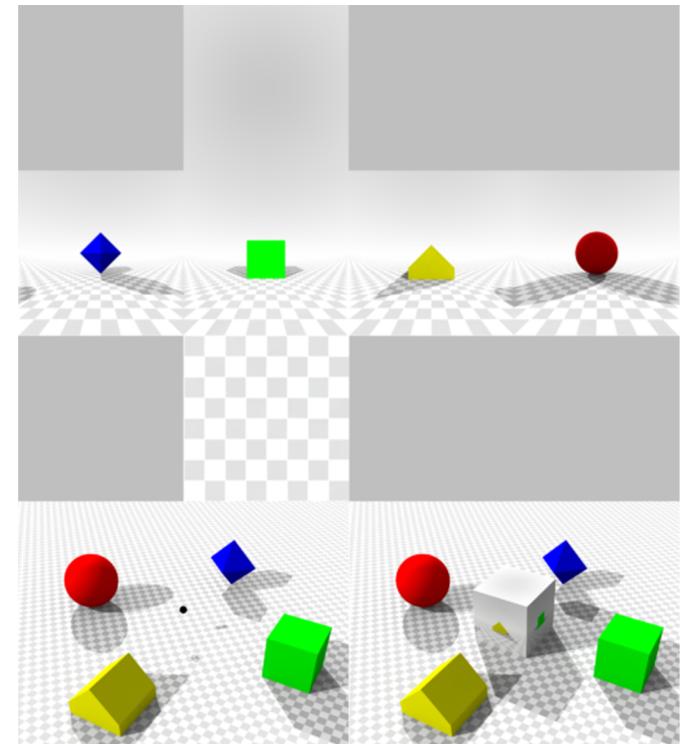
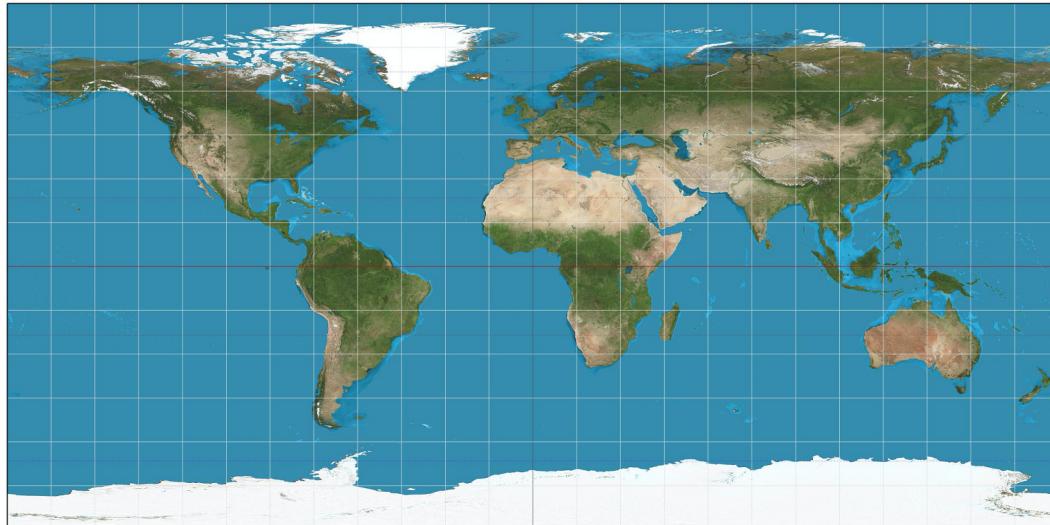


# MPEG-DASH (mp4; H.264)

- Dynamic Adaptive Streaming over HTTP (DASH)
- Designed for delivery over IP using existing HTTP protocol
- Adaptive bitrate - content split into many parts and stored at different bitrates
- Automatically selects from the alternatives the next segment to download and play based on current network conditions.
- Codec-agnostic, which means it can use content encoded with any coding format, such as H.265, H.264, VP9, etc
- CMAF / HLS - will allow over IOS

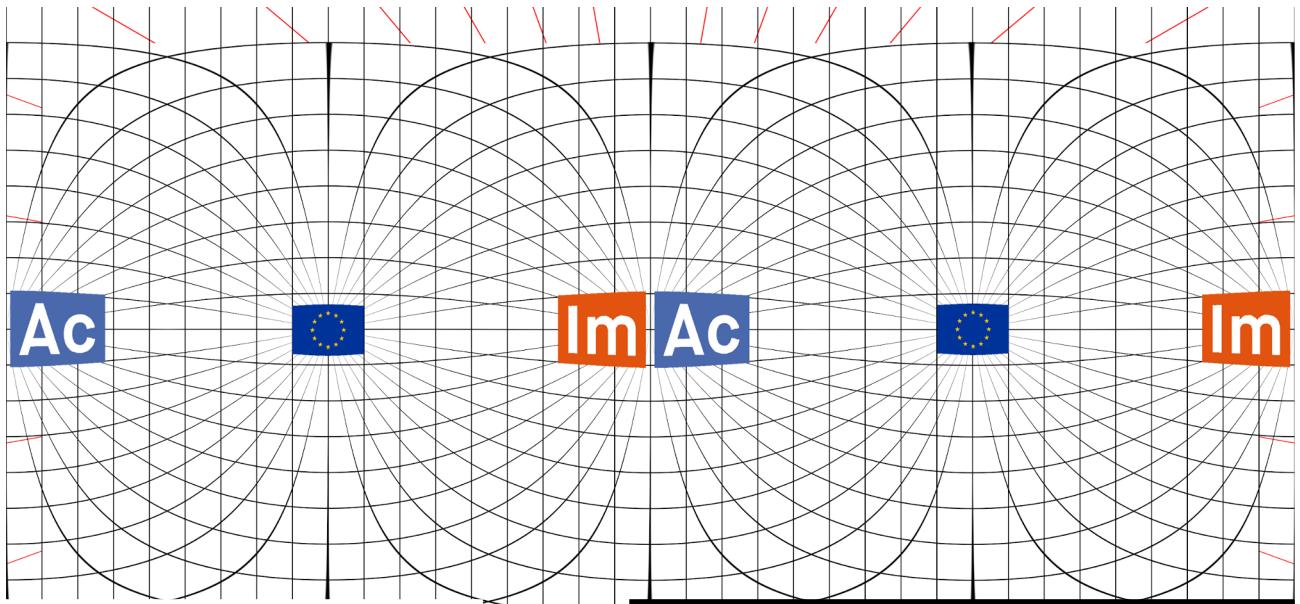
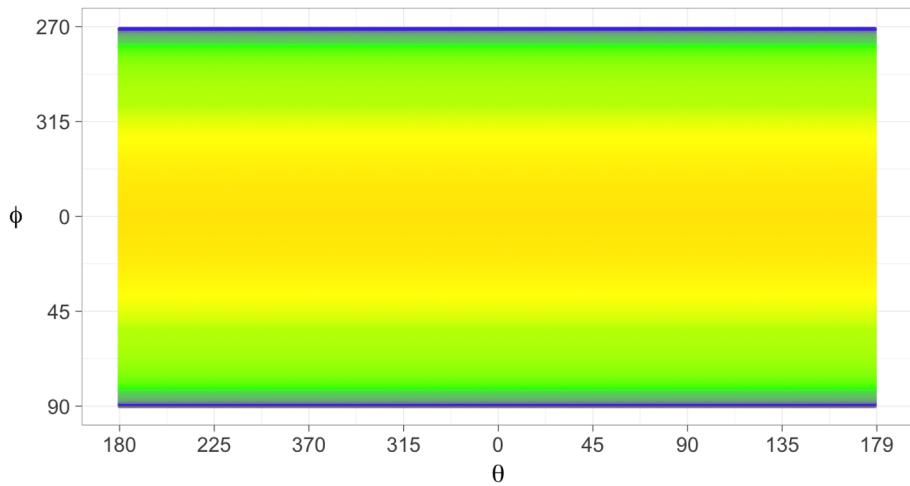
# 360 Video format

- There are two popular formats.
  - Equirectangular
  - Cubemap



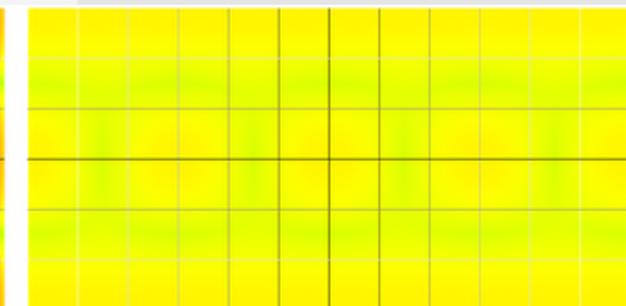
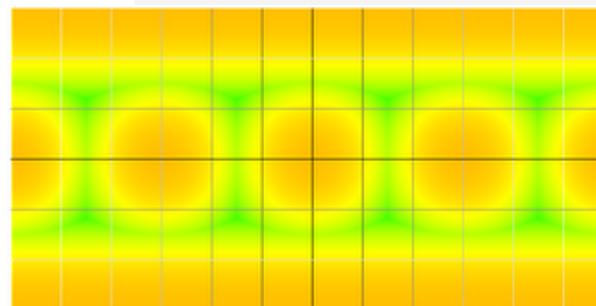
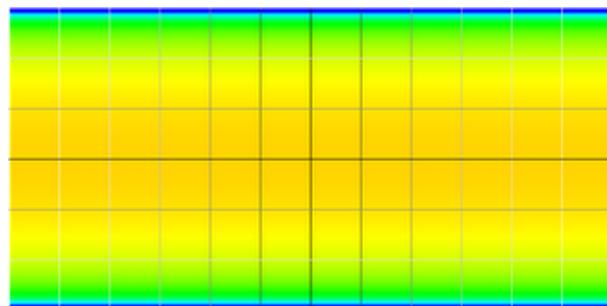
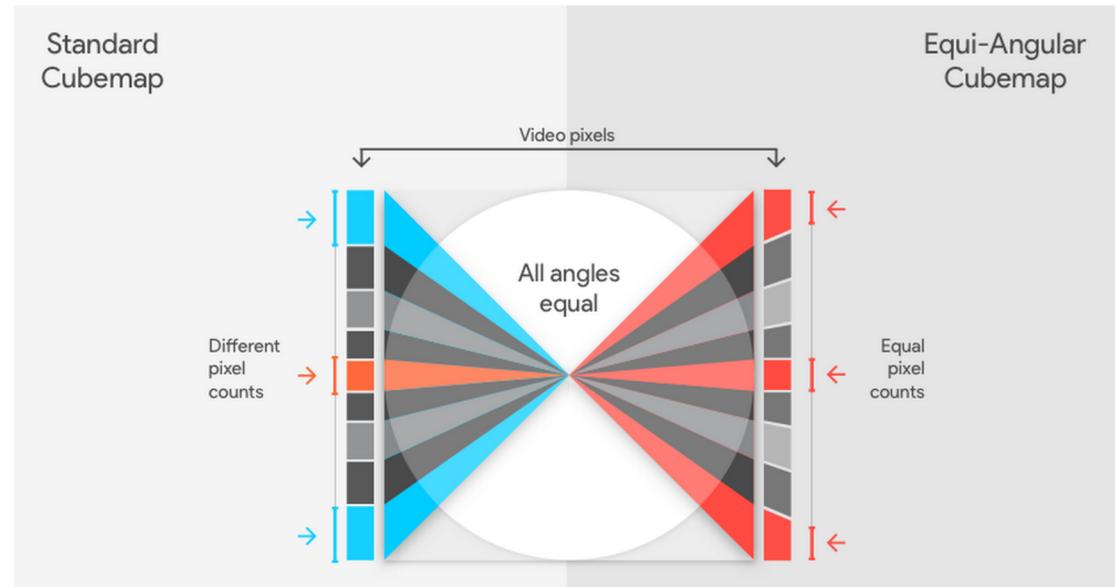
# 360 Video format

- Equirectangular



# 360 Video format

- The Equi-Angular Cubemap (EAC)



Equirectangular Projection (left), Standard Cubemap (middle), Equi-Angular Cubemap (right)

# 360 Video format



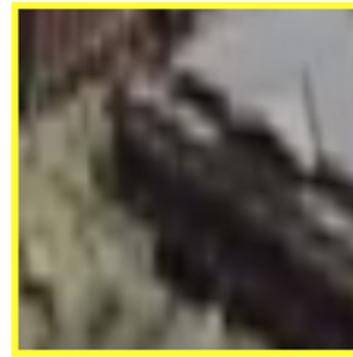
Video frame as viewed in VR headset



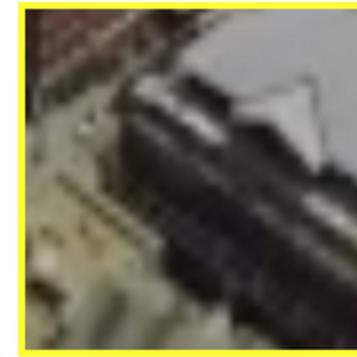
Equirectangular



EAC



Equirectangular



EAC

Source: [www.blog.google/products/google-ar-vr/bringing-pixels-front-and-center-vr-video/](http://www.blog.google/products/google-ar-vr/bringing-pixels-front-and-center-vr-video/)

# Subtitle Distribution

- Options for subtitle distribution:
  - XML-based (TTML)
  - Plain-text (SRT, WebVTT)
  - Embedded in video (CEA-608, CEA-708)
  - Burned into video ('open captions')

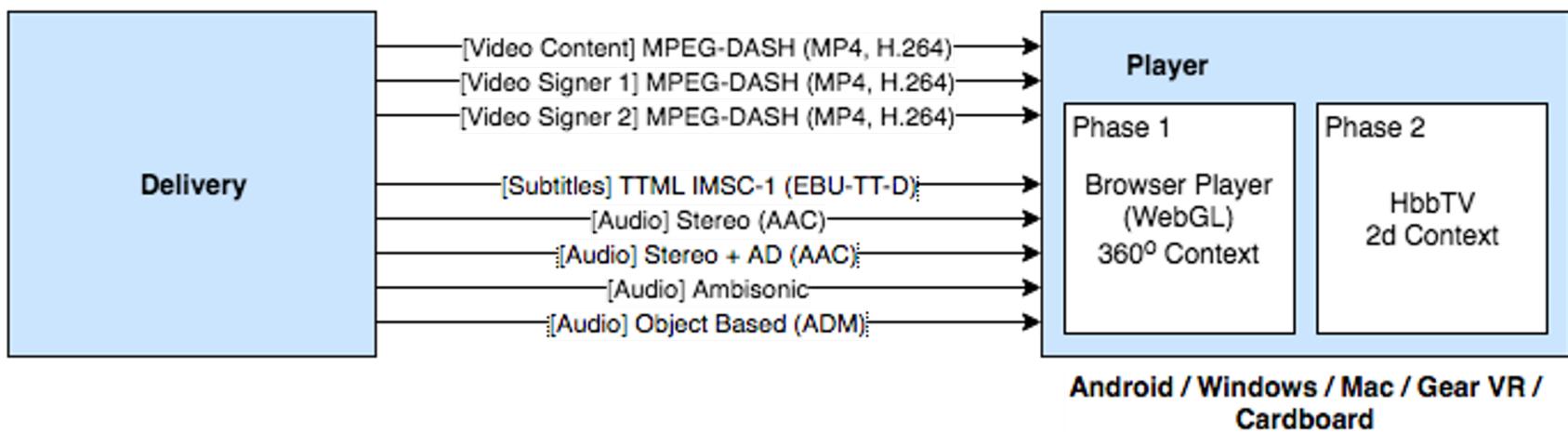
# EBU TTML

- W3C Timed Text working group
- Encompasses all functionalities from existing timed text formats
- Two TTML profiles:
  - EBU-TT, for archiving and as interchange format.
  - EBU-TT-D, which is specifically meant for distribution over IP-based networks.

```
<body>
  <div>
    <p begin="00:00:00.000" end="00:00:02.000">
      This is a subtitle<br/>
      on two lines
    </p>
  </div>
</body>
```

# Delivery

## Delivery



ImAc Player

# ImAc Player - Features

## Use of web technology

- Universal support (cross -device, -browser, -network support)
- No need for installations / updates at the client side

## Supported media formats:

- Traditional 2D and 360º video
- Traditional 2D and spatial audio (Ambisonics)
- Internet Media Subtitles and Captions (IMSC) subtitles (subset of TTML, W3C standard)

## Supported consumption devices (interaction modalities):

- PC and laptops
- Smart TVs
- Mobile devices (tablets, smartphones...)
- VR devices (Head Mounted Displays, HMDs)



# ImAc Player – ST Features

## Styling effects for speakers' representation:

- Color, formats, voice-off

## Integration of Non-Speech Info:

- Examples: test descriptions, emojis
  - Demos: [Contact mario.montaged@i2cat if interest](mailto:mario.montaged@i2cat)

## Personalized presentation:

- Language, Size, Outline and Position
  - Demos: <https://imac.gpac-licensing.com/player/>

## Easy-to-Read Subtitles:

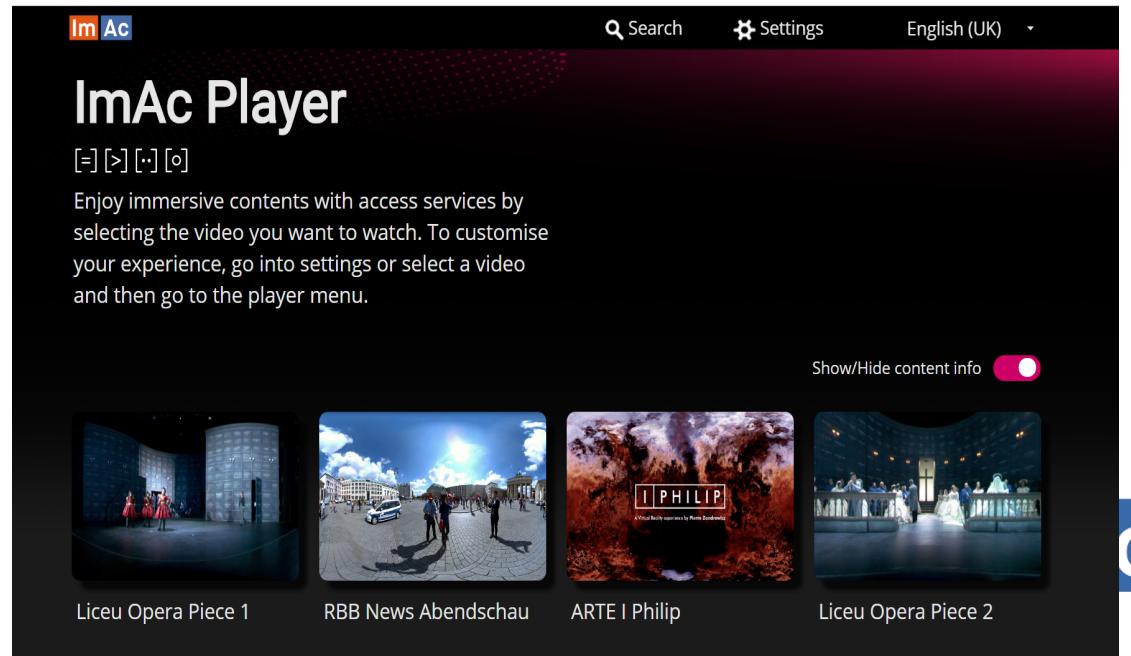
- Validated simpler text structures.
  - Demos: [Contact mario.montaged@i2cat if interest](mailto:mario.montaged@i2cat)

## Responsive Subtitles *[Explained later by Chris Hughes]*

# ImAc Player - UI

ImAc portal. Landpage for:

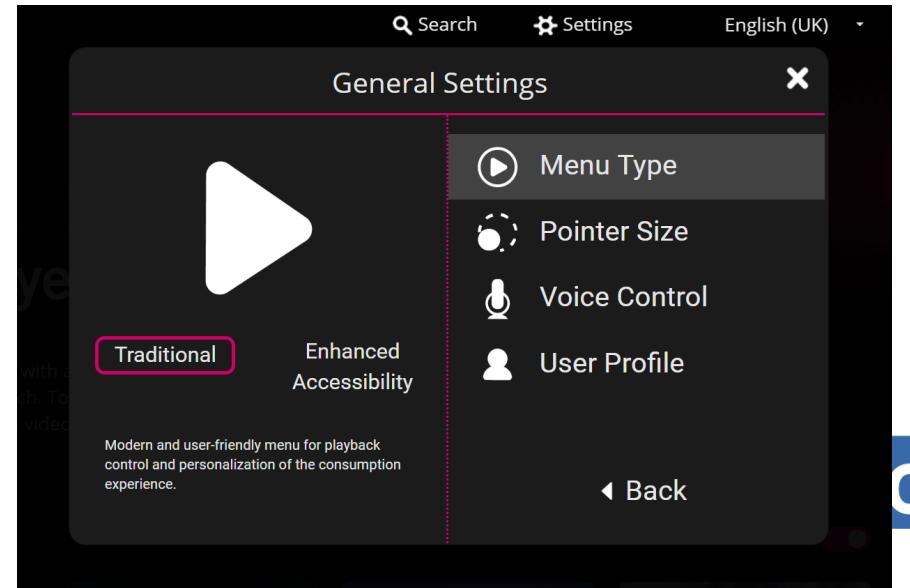
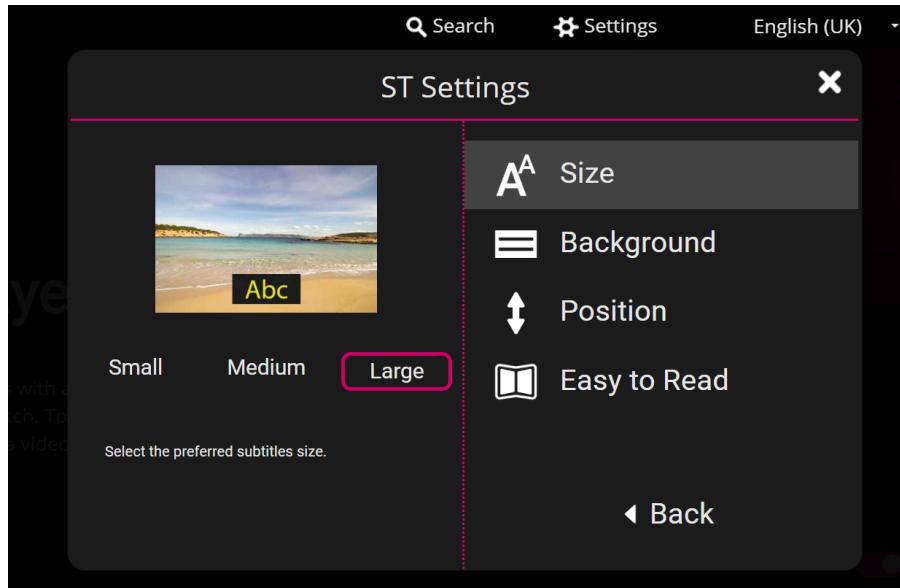
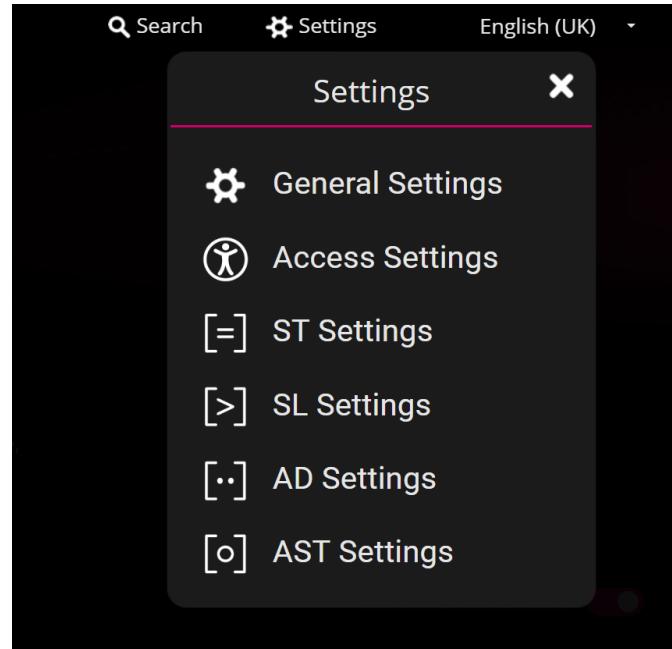
- Language selection
- Catalogue of videos, indicating their language, cover and available access services ( [=] [>] [..] [o] )
- Search and filtering
- Video Selection



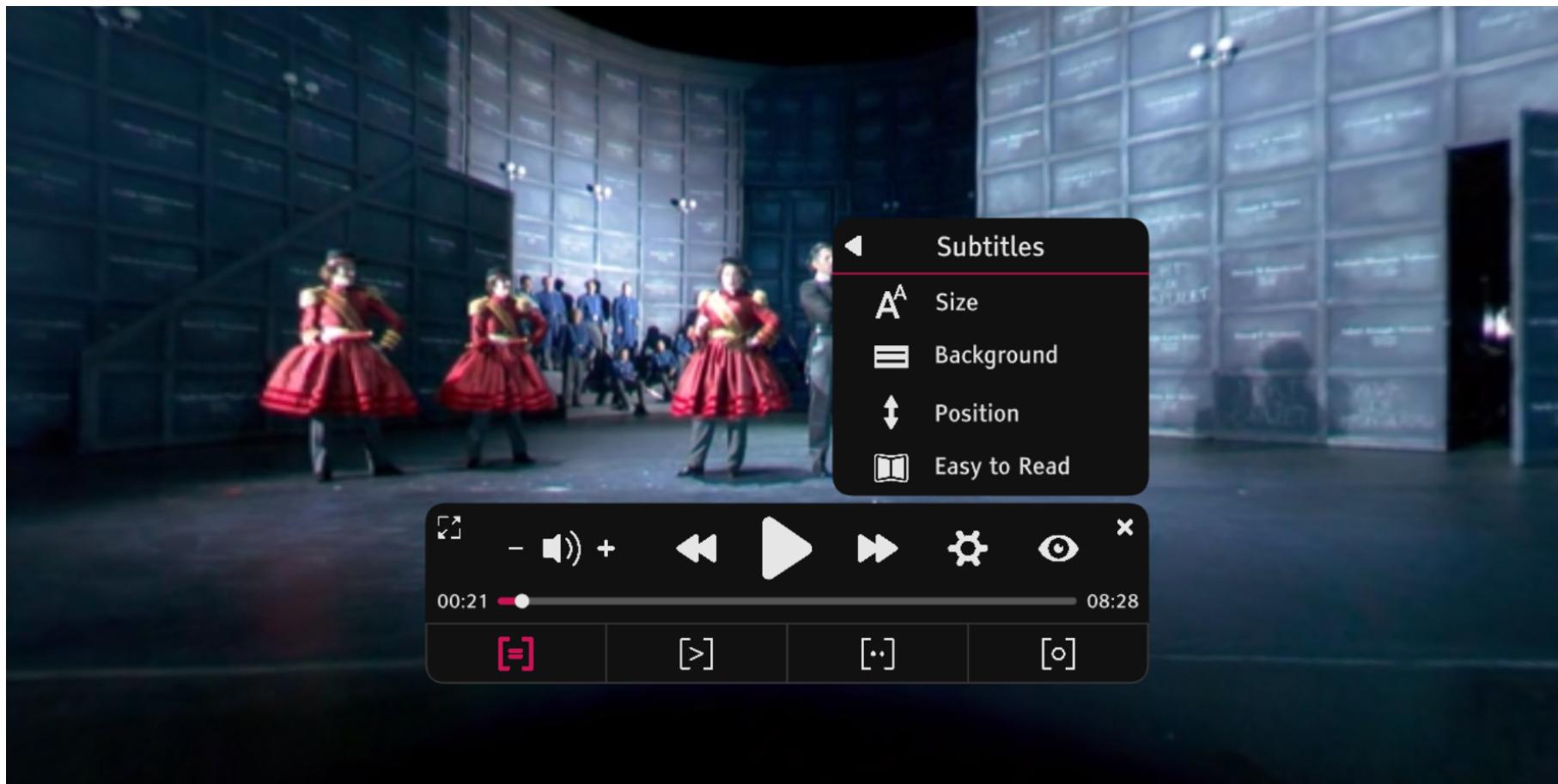
# ImAc Player - UI

ImAc portal. Landpage for:

- Initial Settings



# ImAc Player - UI



# ImAc Player – ST Indicators



# ImAc Player – ST Indicators



# ImAc Player – Multi-Screen



# ImAc Production Tools

# ImAc access service production tools

- **Web subtitle editor (Web ST Editor)** used by subtitlers
- Web audio description editor (Web AD Editor) used by audio describers
- Web sign language editor (Web SL Editor) used by signers
- **Accessibility Content Manager (ACM)** used by broadcasters for:
  - Assigning the production of 360° access services to producers such as **subtitlers**, audio describers or signers (usually external service providers).
  - Sending the LQ 360° video file to the producer so they can carry out the access service production with the editors.
  - Receiving the access service files from the producers and verifying that they are correct.
  - Cataloguing the verified files so they can be used anytime that the 360° programme is broadcasted or published.
  - Automatic background processes (transcodings, renderings, notifications, ...)

# Aproach

- Web ST Editor and ACM embrace:
  - the 360° media
  - the user requirements
  - TTML-based formats (extension of IMSC profile)
  - interconnection with existing systems
  - not losing its backward compatibility (import and export from/to other formats: EBU-TT-D)

### INFO

USERNAME:	test
FILE ID:	346
FILE CREATION DATE:	2019-06-19 11:08:46
ASSET ID:	374
ASSET TITLE:	Holy Land 1
PROGRAMME ID:	
LANGUAGE:	Catalan
VIDEO TYPE:	360
VIDEO SIZE:	51.93 MB
VIDEO DURATION:	00:05:05:13

### SETTINGS

GENERAL SETTINGS

REGIONS: R1 R2 R3 +

CHARACTERS:

Holy Land

### SUBTITLE LIST

1	00:00:00:05	00:00:03:05	Holy Land
---	-------------	-------------	-----------

### VIDEO

00:00:04:23 **GO**

Control buttons:

Waveform:

Control buttons:

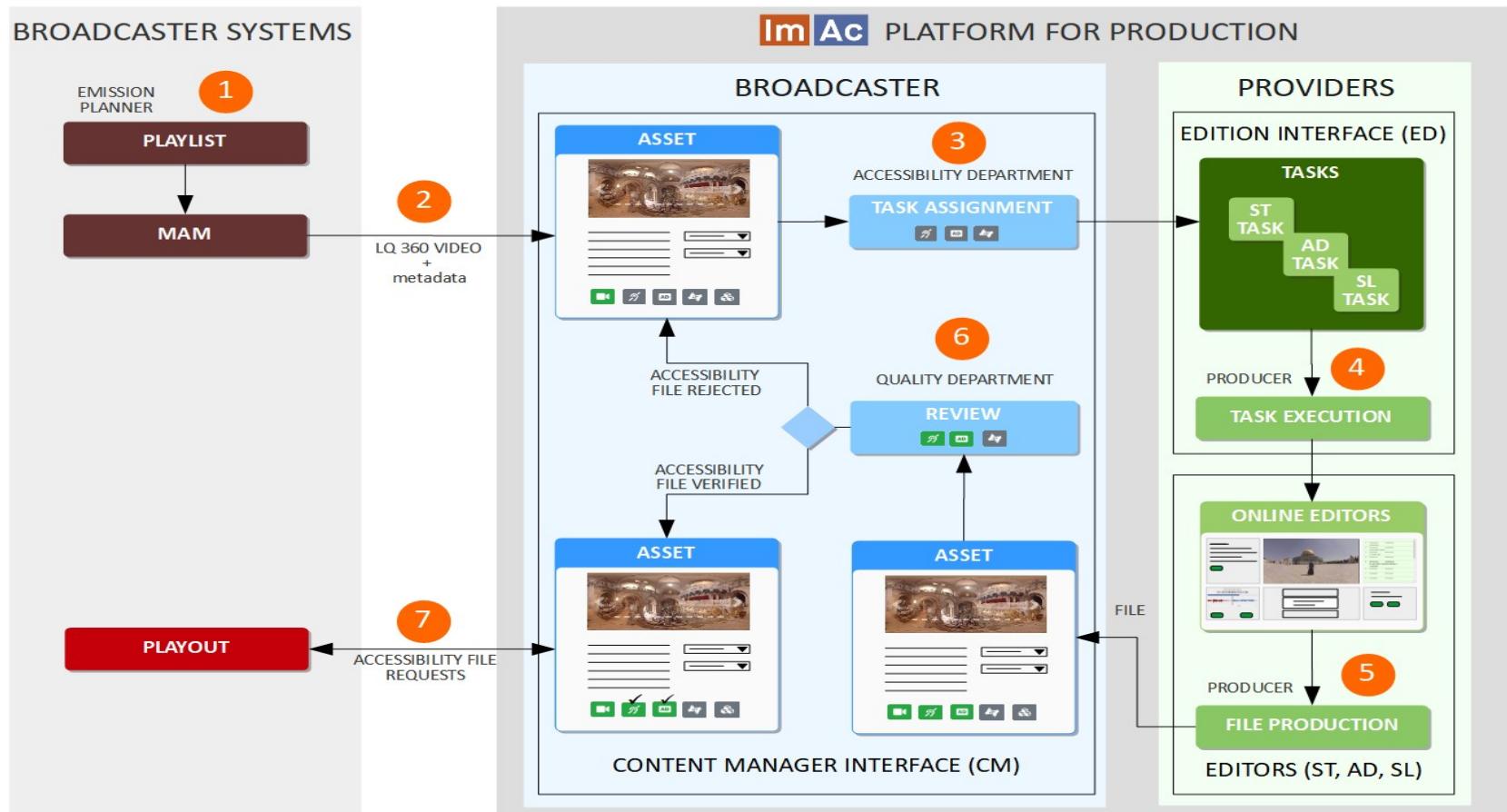
1 /1

R1 R2 R3		Voice Over	+
00:00:00:05	Holy Land	FoV Angle 0.0/0.0	-
00:00:03:05		Speaker's Location 0 0	
00:00:03:00	66 40.00	Saved Angle 0/0	+

### ACTIONS

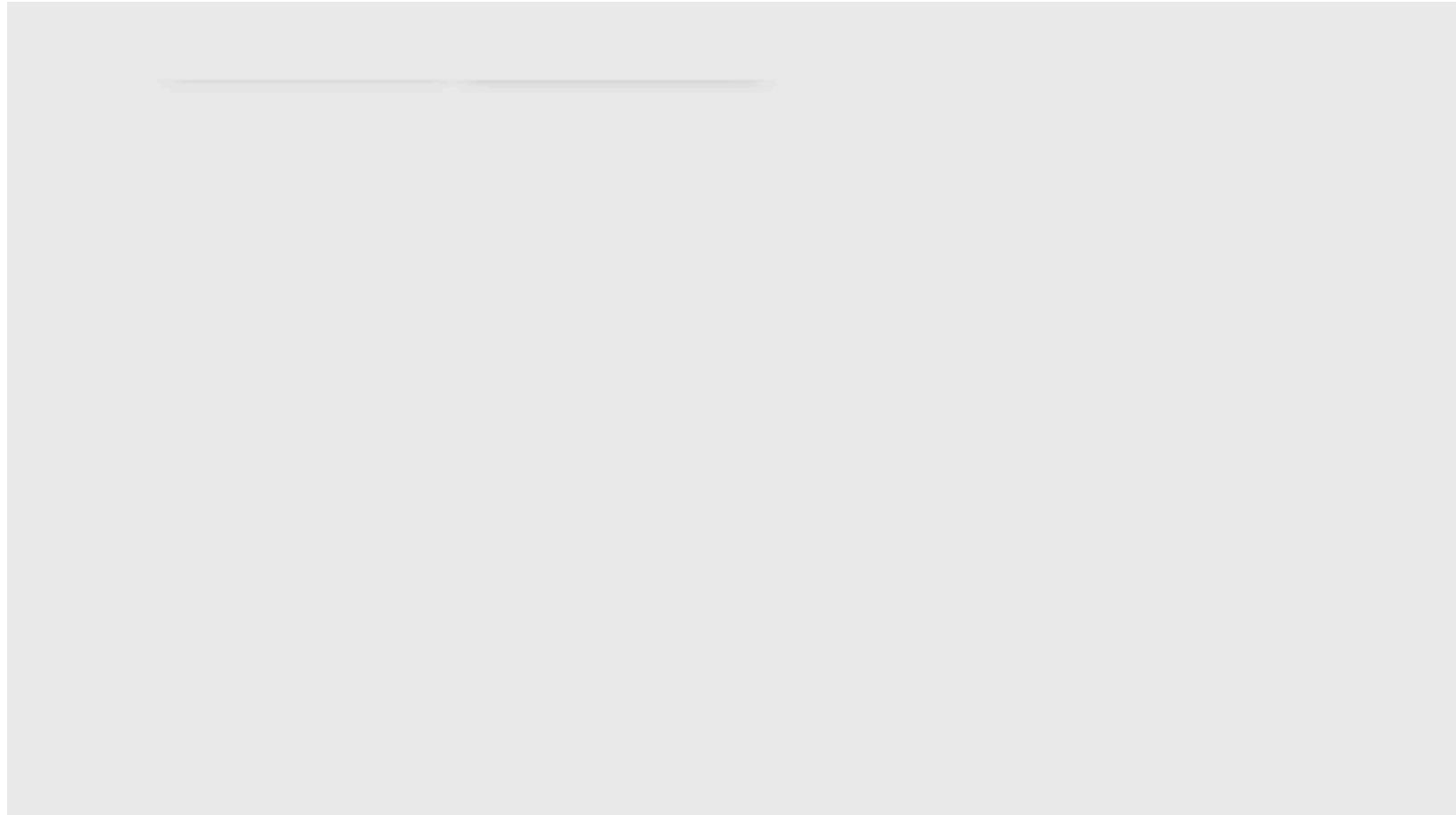
SAVE	AUTO SAVE
FORCED PREVIEW MODE	FREE PREVIEW MODE
CHECK TCS	TC SHIFT
Find	Replace
TEXT FILE IMPORT	

# ImAc production workflow



# Imac Demo Video

- <https://www.youtube.com/watch?v=6U2CUfrijHU>



# Try ImAc

- <https://rbburl.de/machtsbarrierefrei>

## ÜBER DEN RBB

STRUKTUR **PROFIL** ZAHLEN UND FAKTEN GESCHICHTE RBB IN DER ARD RUNDFUNKBEITRAG GREMIEN **INNOVATIONSPROJEKTE**

Profil

**DER RBB MACHT'S**



Kultur statt Corona  
**DER RBB MACHT'S - barrierefrei**

Museen und andere kulturelle Einrichtungen sind geschlossen, das Reisen ist verboten. Wir sorgen trotzdem für vielfältige Erlebnisse. Zu jeder Zeit - von zu Hause aus. Denn **DER RBB MACHT'S** möglich und füllt Lücken, die Corona und Kontaktsperrre in das kulturelle und gesellschaftliche Leben reißen.

Wir laden Sie ein zu unseren 360°-Touren durch aktuelle Ausstellungen und ferne Länder. Und die stehen auch barrierefrei zur Verfügung - **dank des EU-Projekts ImAc**, an dem die rbb Innovationsprojekte beteiligt waren.



# References

- Hughes, CJ , Orero, P and Rai, S 2020, 'Towards a user specification for immersive audio description' , in: Innovation in Audio Description Research , The IATIS Yearbook , Routledge.
- Hughes, CJ , Montagud, M and tho Pesch, P 2019, Disruptive approaches for subtitling in immersive environments , in: TVX 2019 - ACM International Conference on Interactive Experiences for Television and Online Video, 5-7 June 2019, Media City, Salford UK.



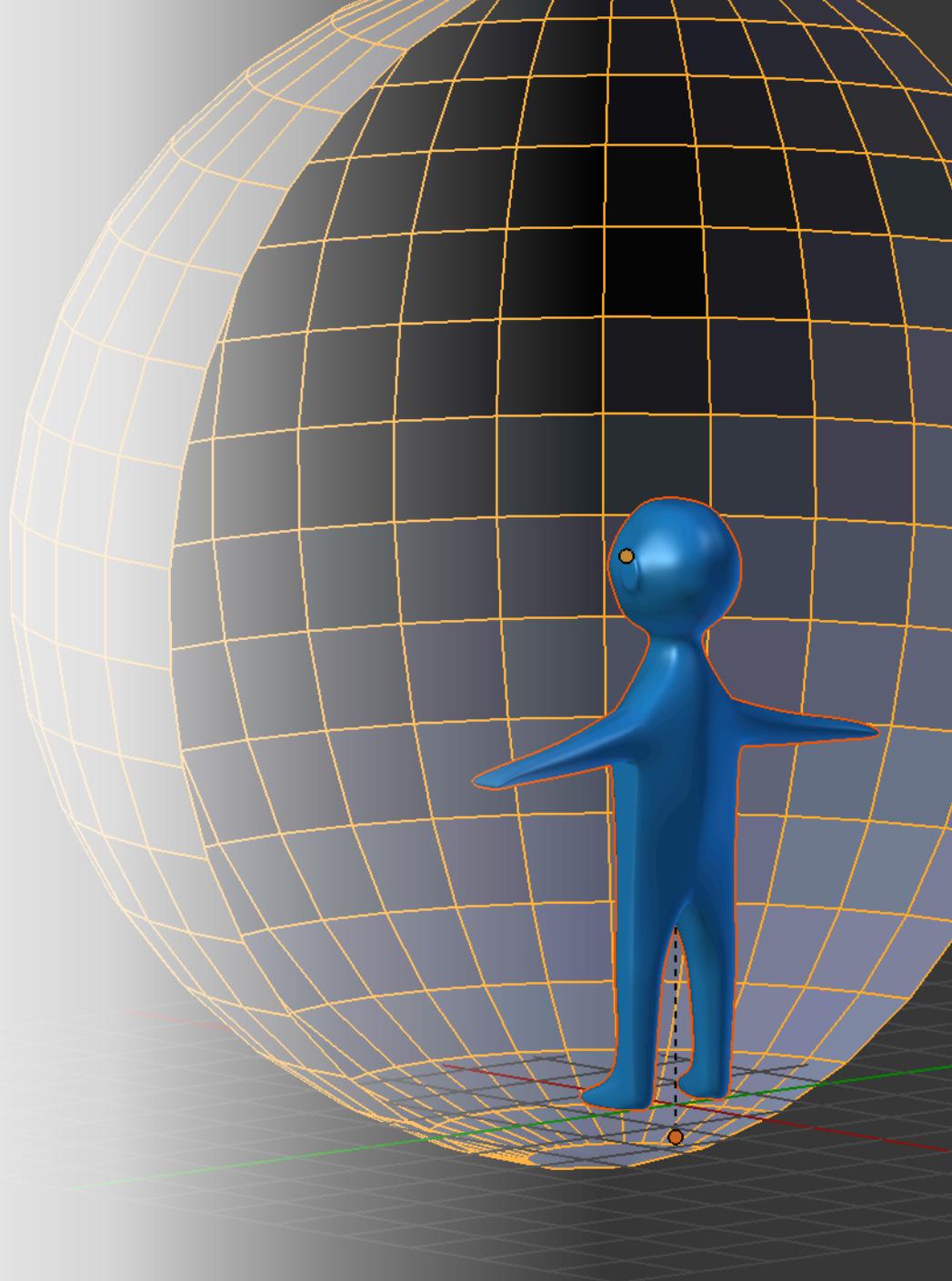
Break



# Next Steps

---

My observations from first large scale EU Project



# Next Steps

- User tests yield limited results unless you can put a working product in front of the user
  - If you offer a paper prototype it is possible to cause confusion
  - Often leads to users saying they prefer what they already have.
- In this area many technologies have a learning step
  - You cannot ask a user to evaluate a prototype while they are learning
- The content is important
  - If the users have no interest in what is being presented it will not be a fair test

# W3C Immersive captions CG

- Lead by Christopher Patnoe (Google)
- The goal of this community group is to determine and publish best practices for access, activation, and display settings for captions with different types of Immersive Media - AR, VR & Games.
- We plan to research current examples, identify best practices, and do research on those ideas over different surfaces: Smartphone AR, AR glasses, VR goggles, etc.
- Where appropriate, we will share our results and discuss opportunities with the TTML WG (recommendations for media online captioning) and the W3C Immersive Web WG (APIs to interact with XR devices and sensors in browsers).
- <https://www.w3.org/community/immersive-captions/>

# Survey

Approaches evaluated	Advantages	Disadvantages
Fixed in Scene (locked in vertical?)	Easy to achieve technically Can be “burned in” as open captions	Still may be difficult to locate depending on where user is looking Switching from one block to the next can be challenging
Fixed in scene, repeated evenly spaced	Easy to achieve technically Can be “burned in” as open captions	Still may be difficult to locate depending on where user is looking Switching from one block to the next can be challenging
Appear in front, then fixed in scene	Always appear in view, but don’t distract after appearing	Could appear in the wrong position while user is moving their head quickly
Fixed, position in scene	Clear context from where the voice is coming from in the scene	Missed if user is looking in a different direction Production cost to position captions properly, not needed in other approaches Must consider depth in scene, or render on top, resulting in a “punched-in” effect
Head-locked	Always in view Simple & easy to understand	Position may be distracting, feel “stuck to your face” Objects fixed to head may cause discomfort Position may occlude content
Head-locked on horizontal axis only	Always in view	May be out of view when looking up/down Fixing in only one direction seems more distracting than follow head
Head-locked with lag, animate into view	Always in view Avoids head-locked discomfort	Animation feel unpredictable, reduce immersion
Head-locked with lag, jump into view	Stays in view	Jumping is somewhat unpredictable, distracting when it happens
Fixed, positioned in scene, then head-locked with lag	Clear context from where the voice is coming from, while staying always staying in view	Production cost to position captions properly, not needed in other approaches

# Survey

Approaches evaluated	Advantages	Disadvantages
Fixed in Scene (locked in vertical?)	Easy to achieve technically Can be “burned in” as open captions	Still may be difficult to locate depending on where user is looking Switching from one block to the next can be challenging
Fixed in scene, repeated evenly spaced	Easy to achieve technically Can be “burned in” as open captions	Still may be difficult to locate depending on where user is looking Switching from one block to the next can be challenging
Appear in front, then fixed in scene	Always appear in view, but don't distract after appearing	Could appear in the wrong position while user is moving their head quickly
Fixed, position in scene	Clear context from where the voice is coming from in the scene	Missed if user is looking in a different direction Production cost to position captions properly, not needed in other approaches Must consider depth in scene, or render on top, resulting in a “punched-in” effect
Head-locked	Always in view Simple & easy to understand	Position may be distracting, feel “stuck to your face” Objects fixed to head may cause discomfort Position may occlude content
Head-locked on horizontal axis only	Always in view	May be out of view when looking up/down Fixing in only one direction seems more distracting than follow head
Head-locked with lag, animate into view	Always in view Avoids head-locked discomfort	Animation feel unpredictable, reduce immersion
Head-locked with lag, jump into view	Stays in view	Jumping is somewhat unpredictable, distracting when it happens
Fixed, positioned in scene, then head-locked with lag	Clear context from where the voice is coming from, while staying always staying in view	Production cost to position captions properly, not needed in other approaches

What does this all mean???

# Hacking Rapid Prototyping

- **What do we think we want?**
- **How do you know without trying?**
- Three rapid development cycles:
  - Hack 1 – Show users something that works for every platform
    - Desktop, tablet, HMD
  - Hack 2 – Add customization options / better views
  - Hack 3 – Realize that Hack 2 doesn't solve the problem and start again.

# Hack 1



# Hack 1

Immersive Subtitle Display Modes (Click to drag view, button to advance captions / or play, select mode)

Play Next Caption Previous Caption Hide Details Head-locked London Bridge Auto Jump

Head-locked

Originally Identified in Matthew Johnston's Survey.

View Angle: 0.53 Subtitle Angle: -2.93 Timecode: 4.65

Subtitle Display Modes (Click to drag view, button to advance captions / or play, select mode)

Previous Caption Hide Details Head-locked London Bridge Auto Jump

- Fixed in Scene (Locked Vertical?)
- Fixed in scene, repeated evenly spaced
- Appear in front, then fixed in scene
- Fixed, position in scene
- ✓ Head-locked**
- Head-locked on horizontal axis only
- Head-locked on vertical axis only
- Head-locked with lag, animate into view
- Head-locked with lag, jump into view
- ImAc Arrow
- ImAc Radar

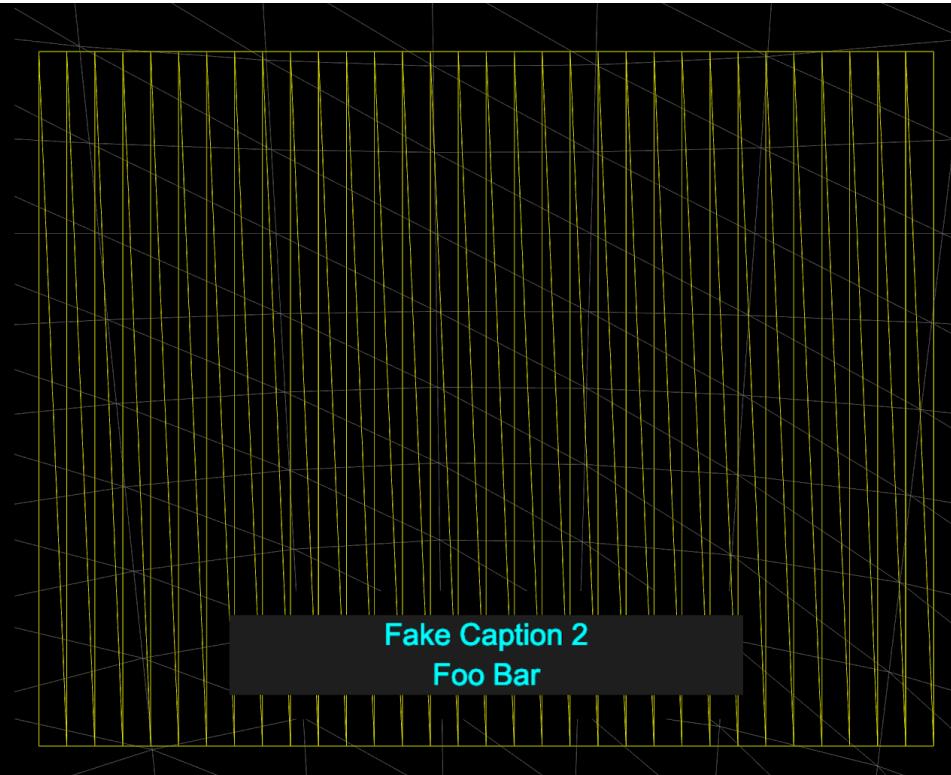
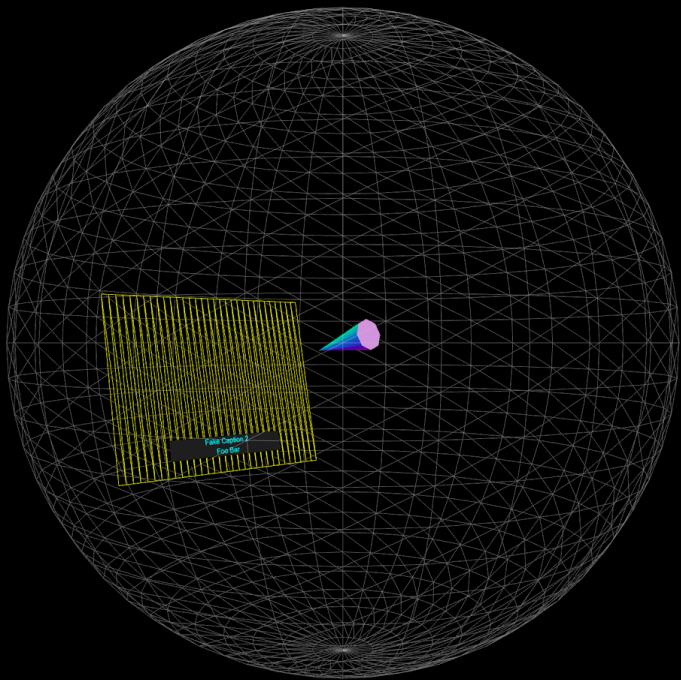
This is an example caption  
People Walking, Talking

[https://www.chxr.org/immersive\\_subs2\\_1](https://www.chxr.org/immersive_subs2_1)

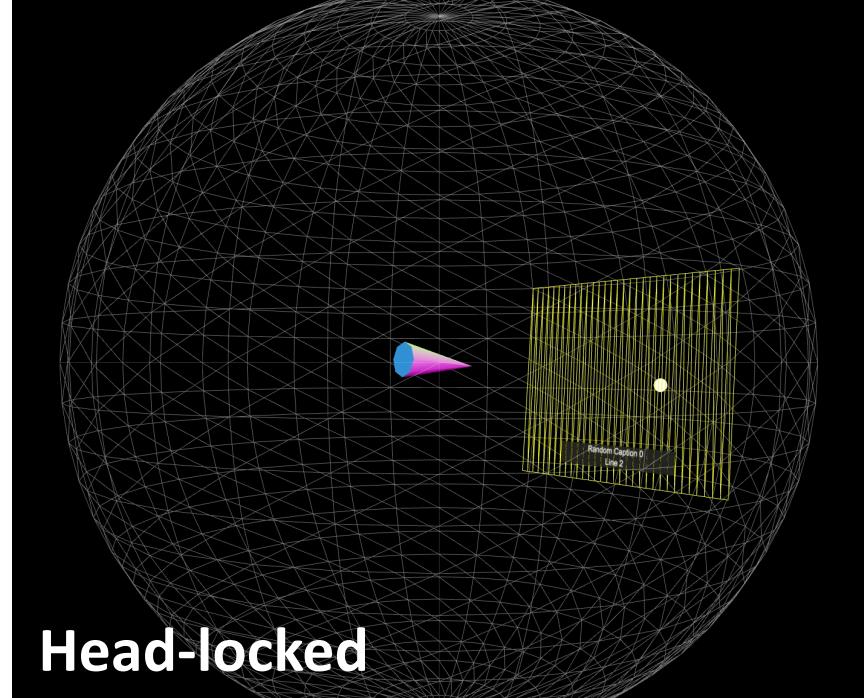
# Hack 1

- Built in 24 hours (Overnight to those in the US)
- Web based – can be distributed by sending a link
- Rights free video clip found on the internet
- JavaScript – it is fast enough
- Three.js graphics library
  - WebVR Polyfill – WebVR/XR for Javascript
- Very badly written, resource heavy and tries to do too many things
  - but it progressed the discussion

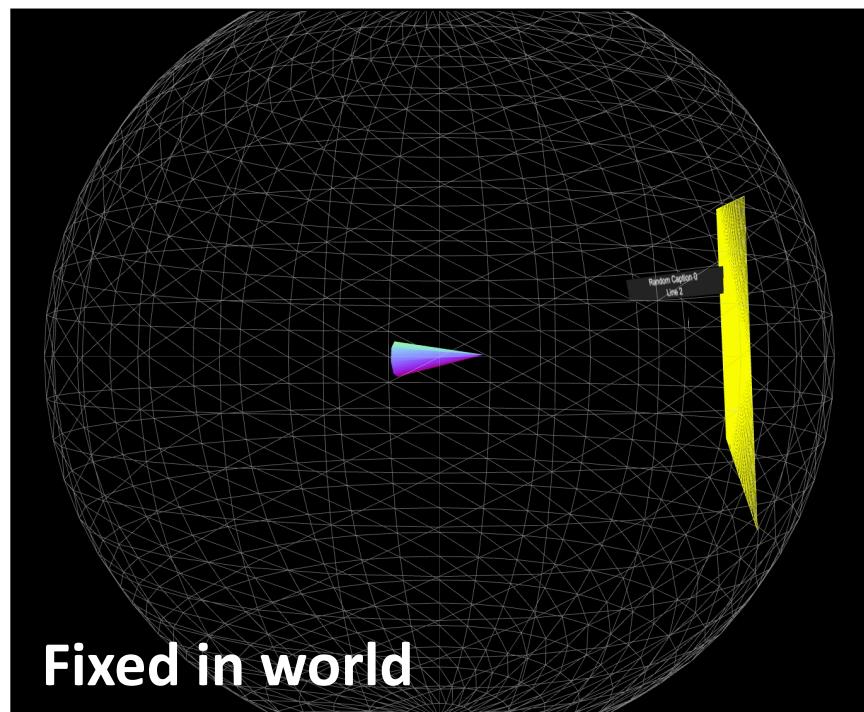
# Hack 2



- **Fixed in Scene, Locked Vertical** - The caption is positioned at the target, but the azimuthal position ( $\phi$ ) is restricted to 0 so that it remains locked to the horizon.
- **Fixed in scene, repeated evenly spaced** - The caption is positioned at the target location then duplicated at  $2\pi/3$  rad ( $120^\circ$ ) intervals around the horizon.
- **Appear in front, then fixed in scene** - The caption is rendered in the centre of the user's current view and remains there until the caption is updated.
- **Fixed, position in scene** - The caption is rendered at the target location.
- **Head-locked** - The caption is rendered in the user's view point and is moved in sync with the user to ensure the caption remains statically attached to the view point.
- **Head-locked on horizontal axis only** - The caption is rendered as *head-locked*, however the azimuthal angle ( $\phi$ ) is restricted to 0, ensuring that the caption is always rendered on the horizon.
- **Head-locked on vertical axis only** - The caption is rendered as head-locked, however the polar position ( $\theta$ ) is locked to the target.
- **Head-locked with lag, animate into view** - The caption is rendered in the *head-locked* position, however as the users' viewpoint changes the caption is pulled back towards the *head-locked* position. An animation loop moves the caption incrementally causing it to smoothly animate into view.
- **Head-locked with lag, jump into view** - This is the same as above, except the animation time is reduced to 0, forcing the caption to jump into the users view.



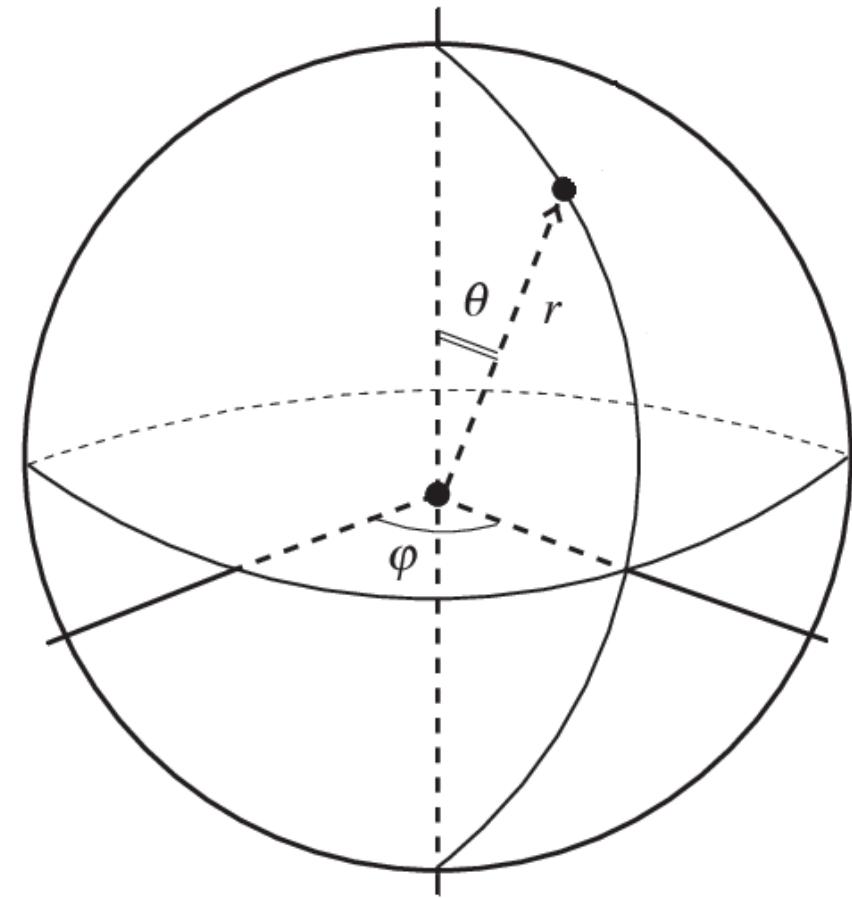
Head-locked



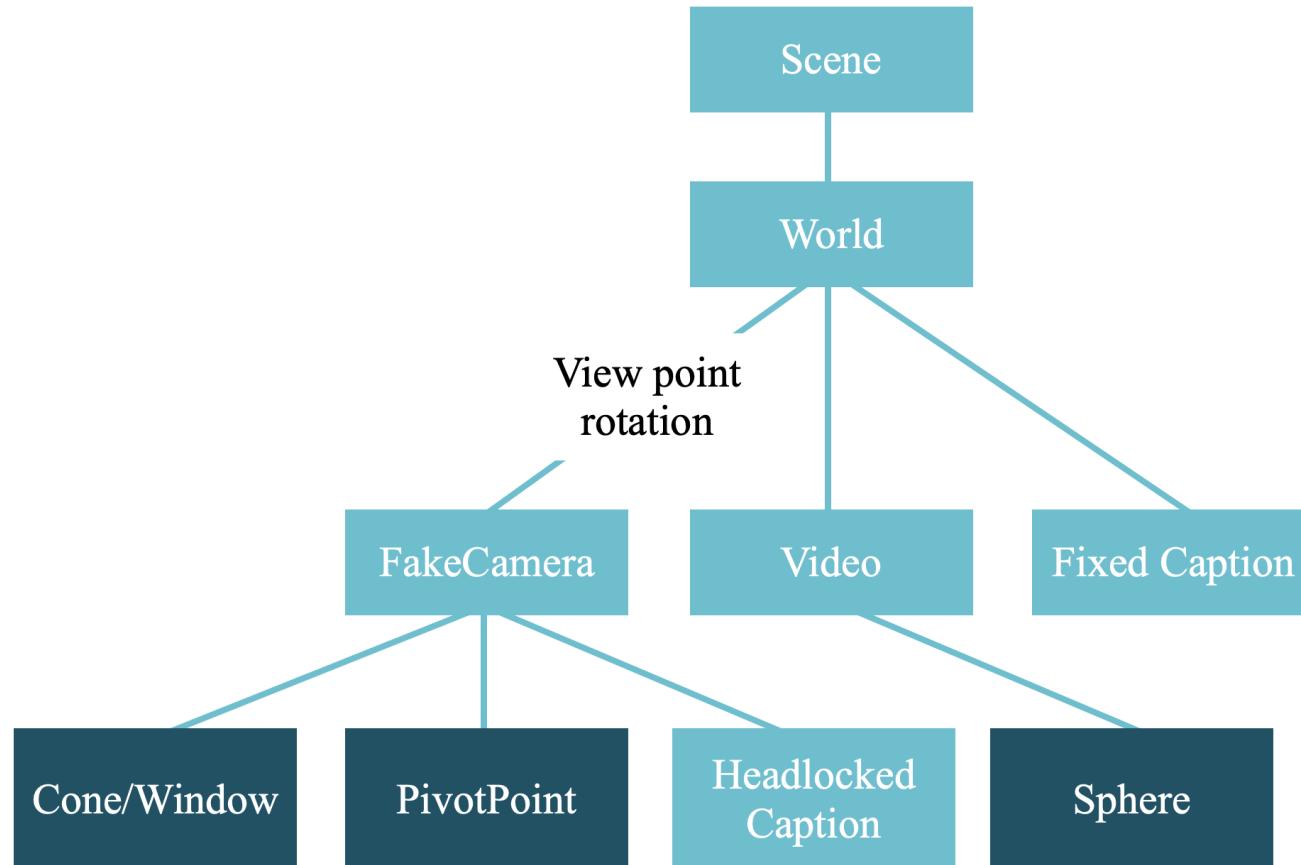
Fixed in world

# Polar Coordinates

- Position the caption anywhere in the scene using a spherical coordinate system
- radial distance ( $r$ )
- polar angle ( $\theta$  - theta) and azimuthal angle ( $\phi$  - phi)
- values are stored in the caption file



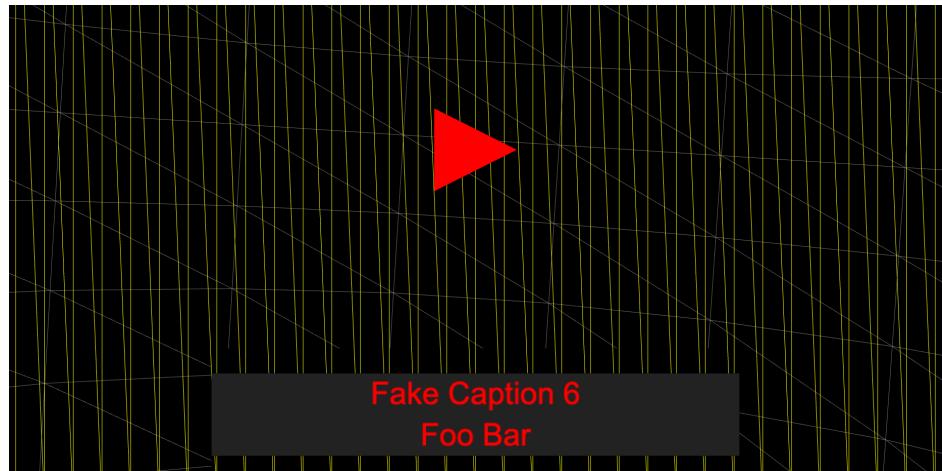
# Scene Graph



- Groups
- Objects

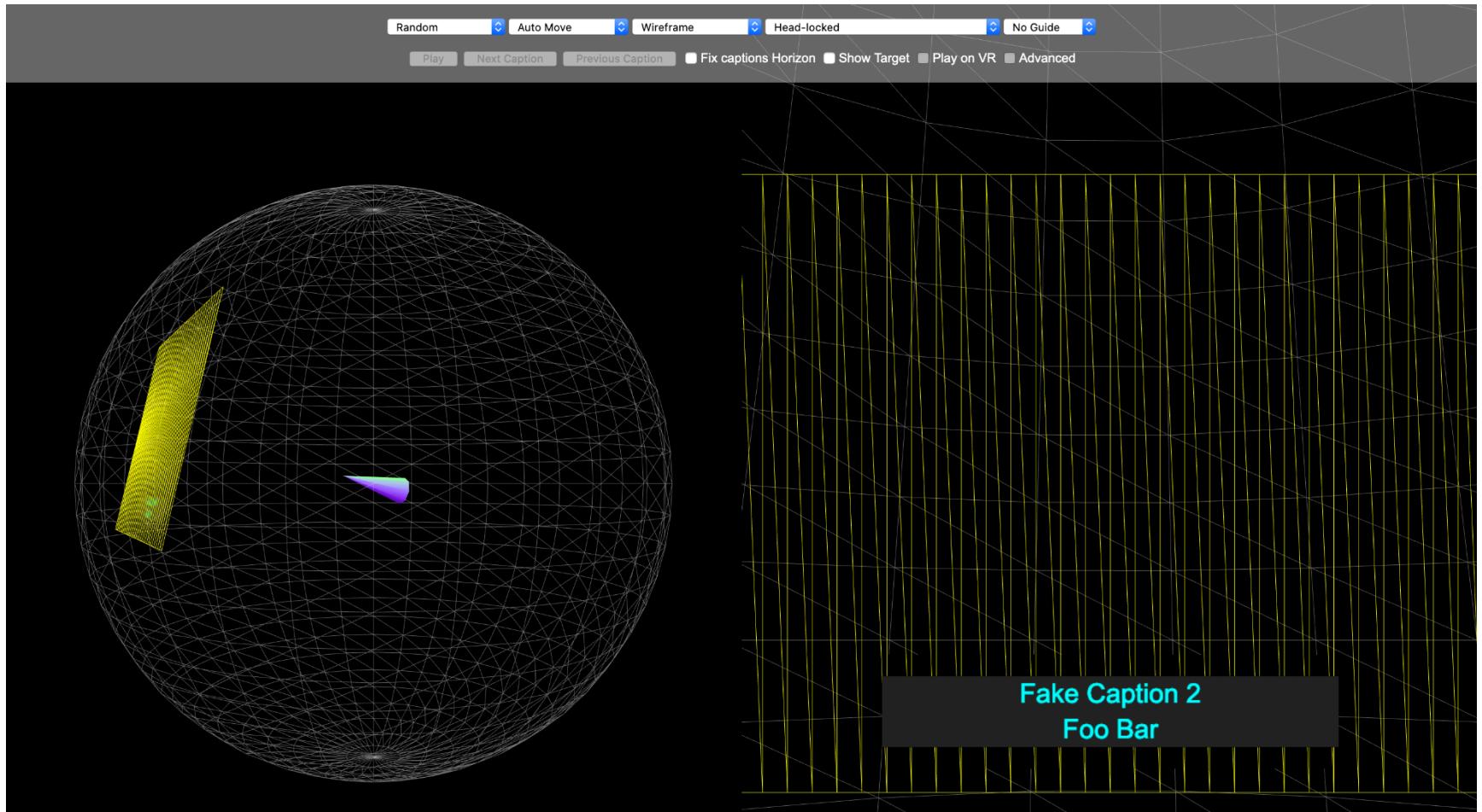
# Guide Modes

- Guide Modes were replicated

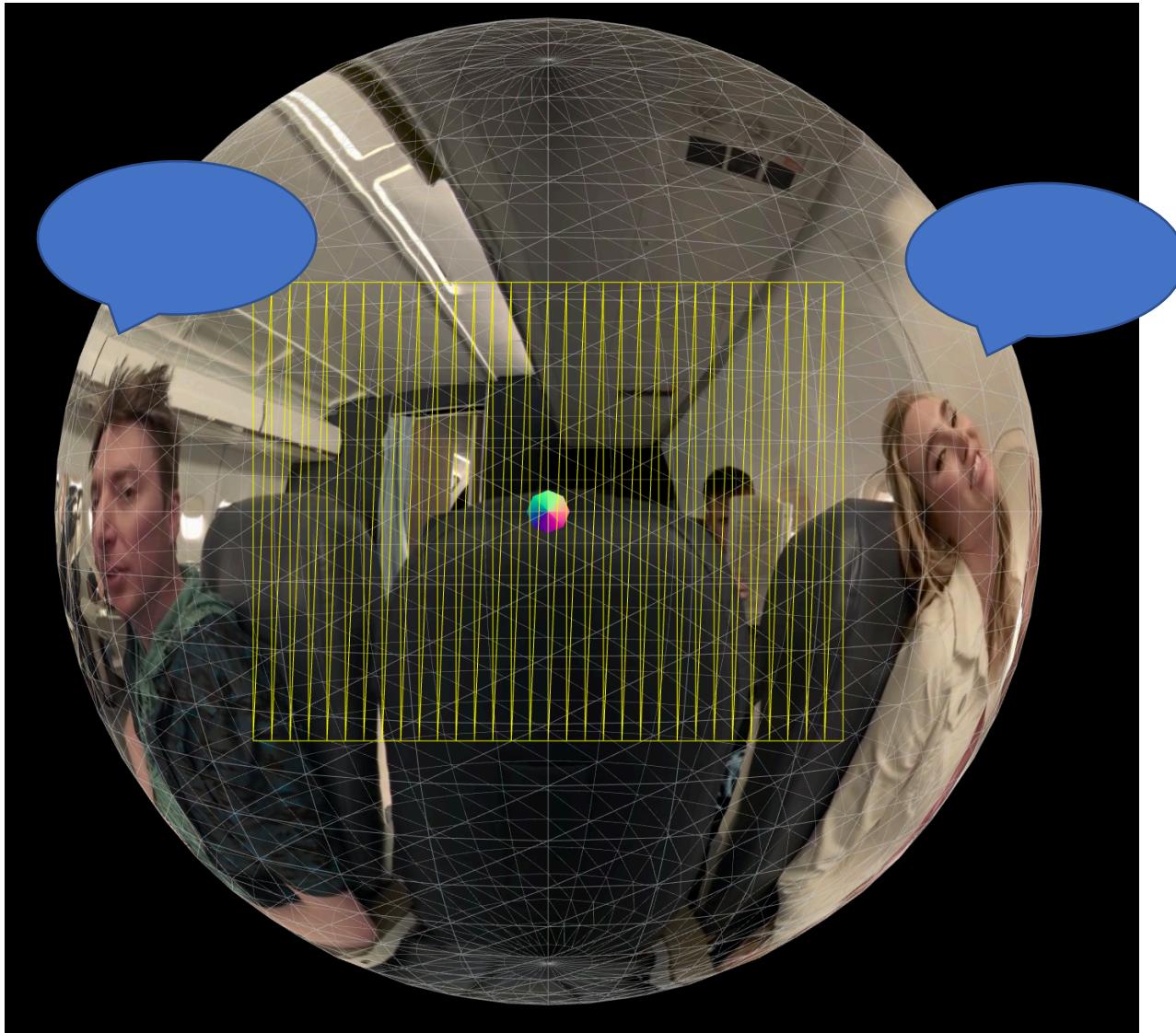


# Try Version 2

- [https://www.chxr.org/immersive\\_subs2\\_2](https://www.chxr.org/immersive_subs2_2)



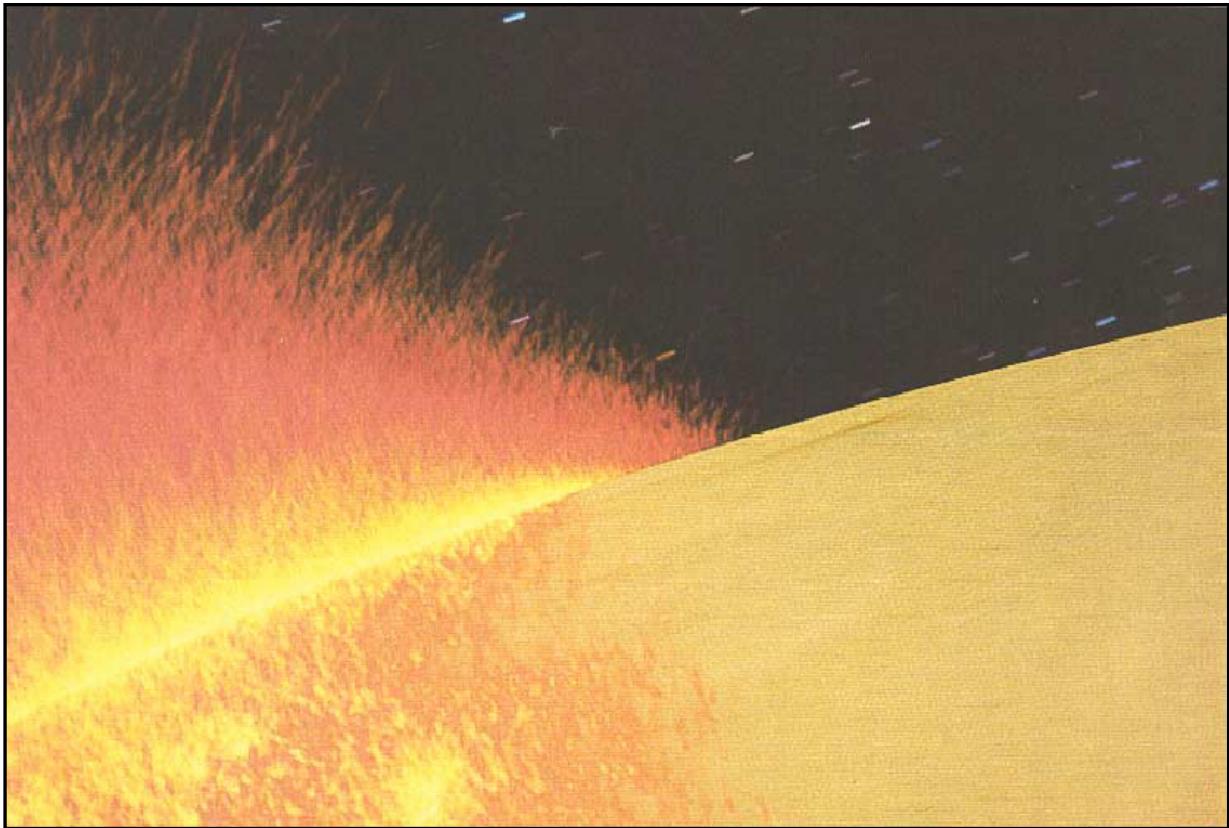
# Hack 2 – Perfect?



# Hack 3

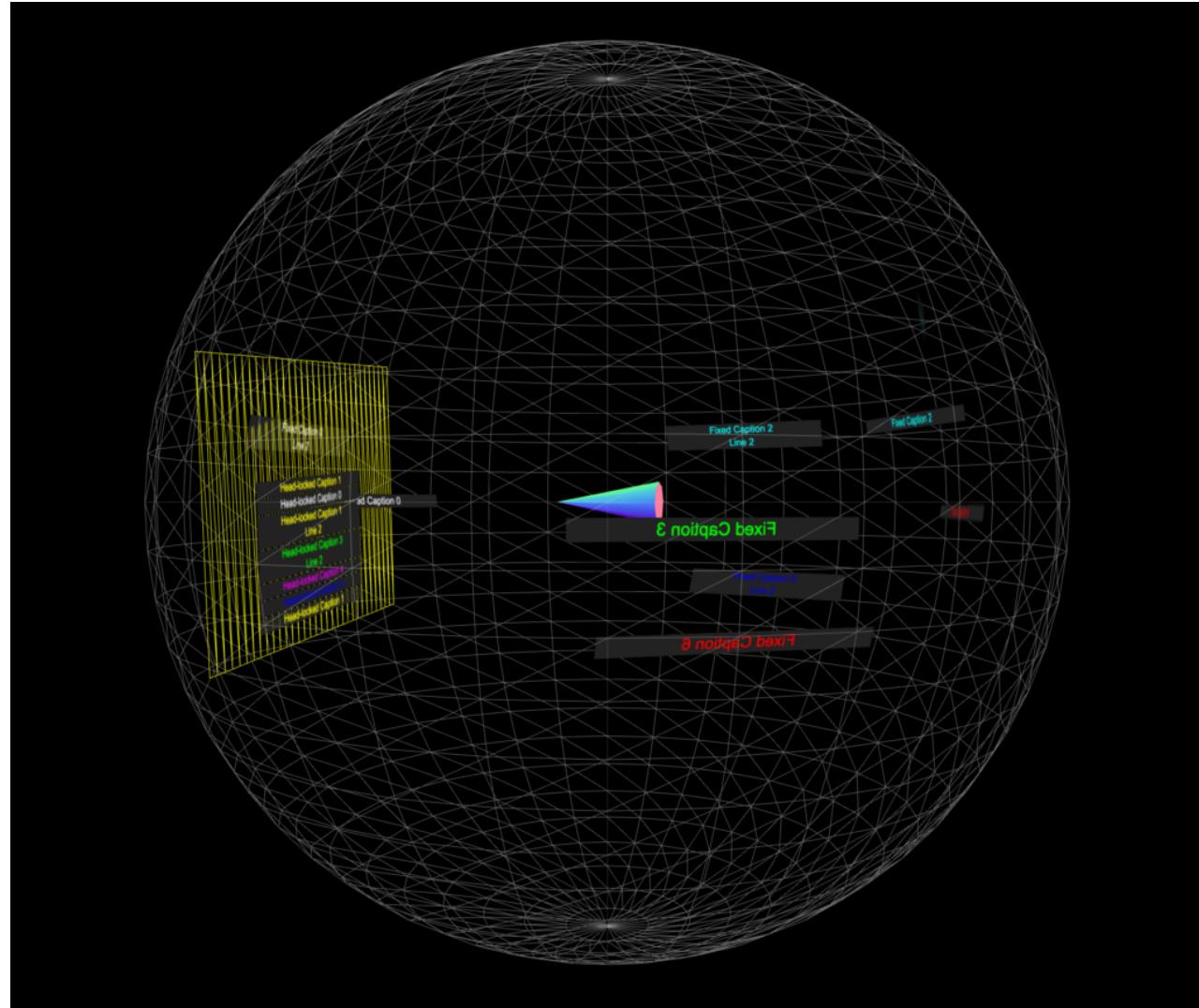
## Particle Systems

- Position
- Velocity
- Color
- Lifetime
- Age
- Shape
- Size
- Transparency



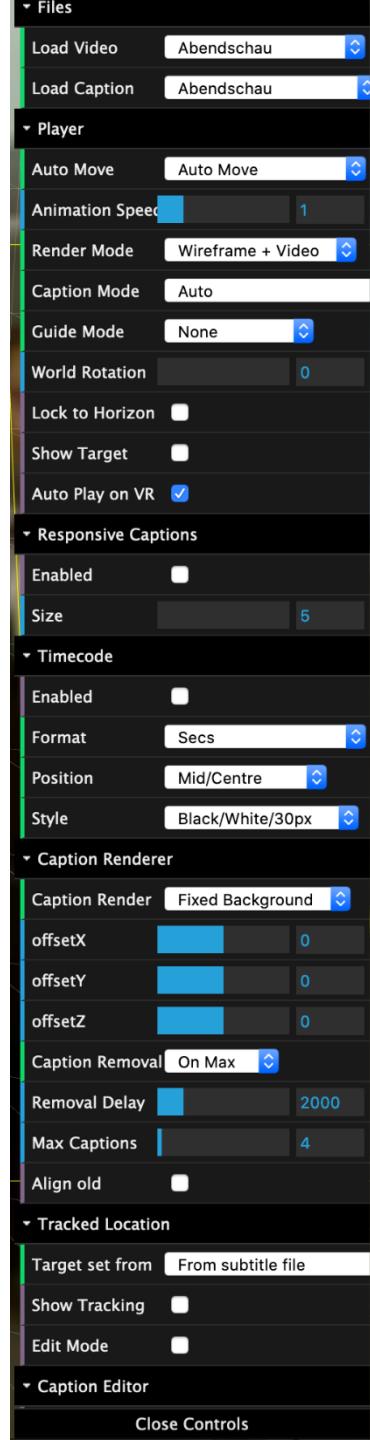
# Captions Object based

- Caption Emitter
- Caption Manager

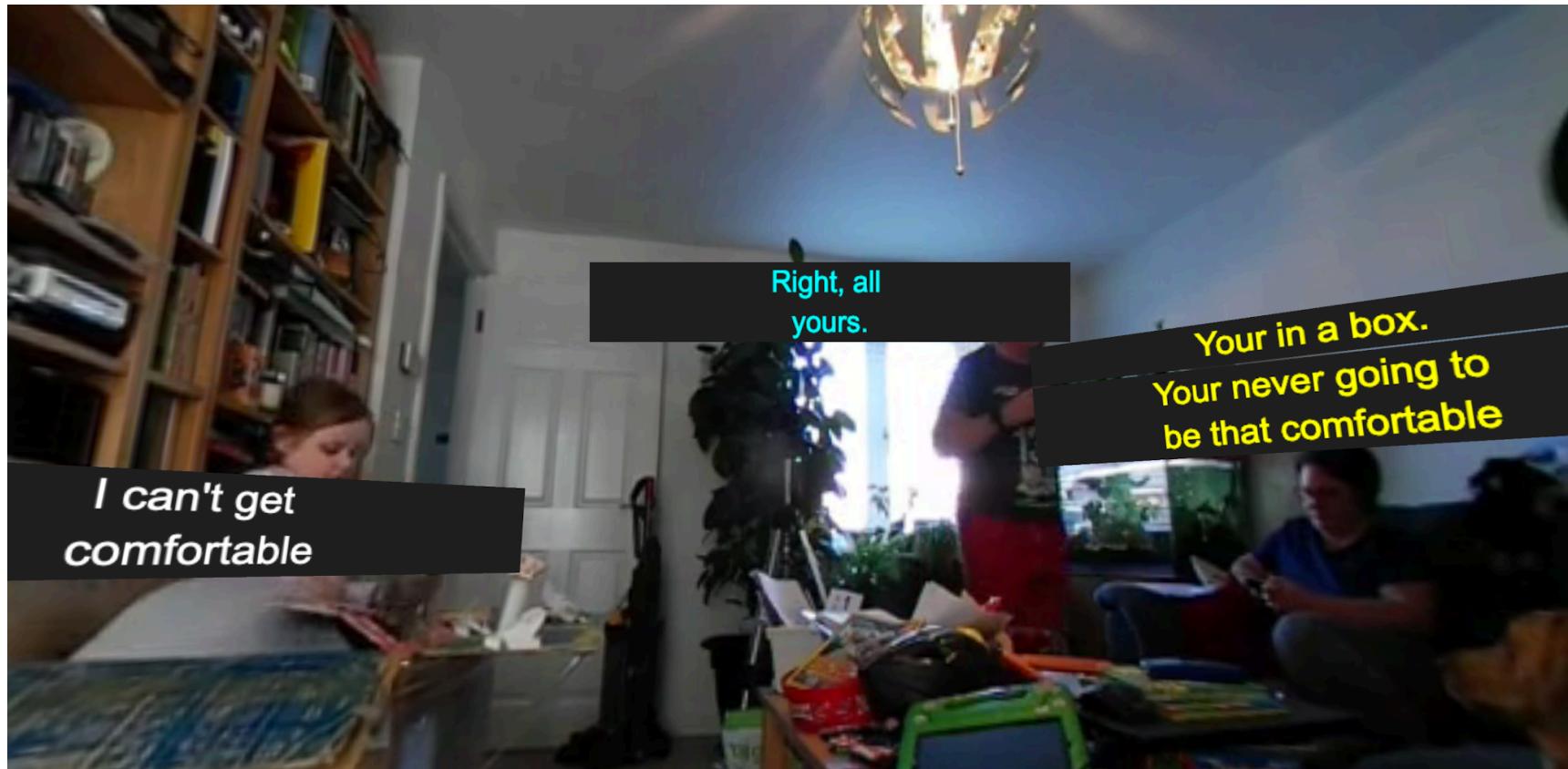


# Experimental Framework

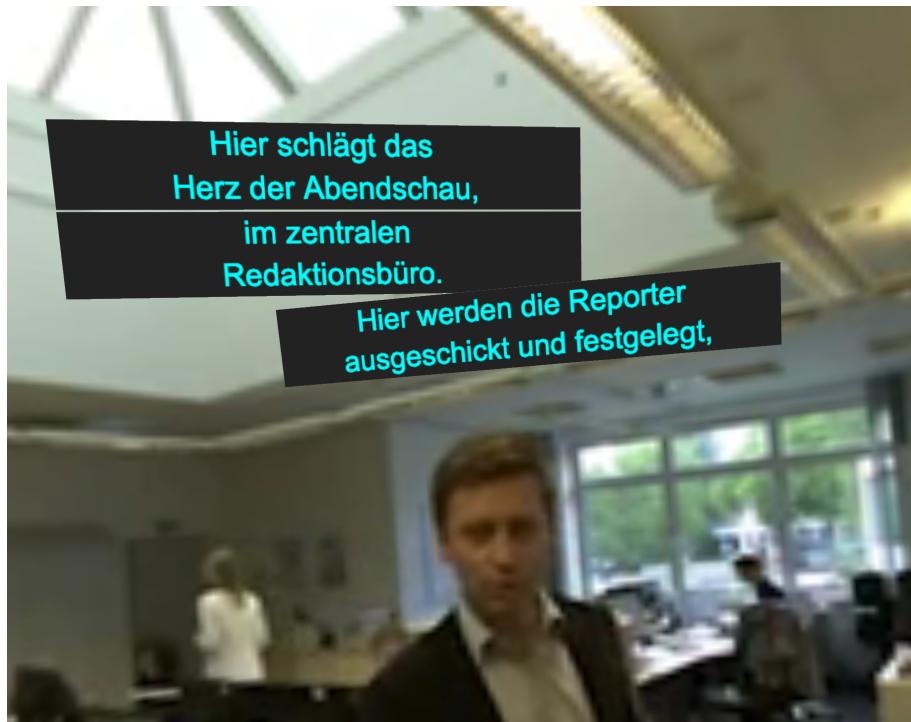
- DASH, HLS, Local Files
- Auto move
- Caption Modes
- Guide Modes
- Responsive Captions
- Timecode
- Custom Renderer
  - Offsets
  - Handle removal
- Location Tracking
- Caption Editor



# Captions fixed to speaker



# Stacking (Collisions)



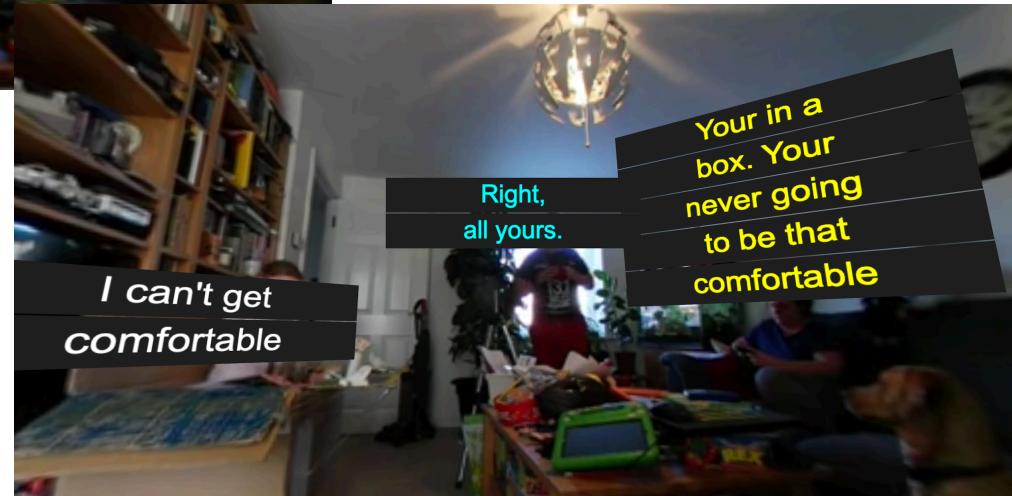
Hier schlägt das  
Herz der Abendschau,  
im zentralen  
Redaktionsbüro.

Hier werden die Reporter  
ausgeschickt und festgelegt,

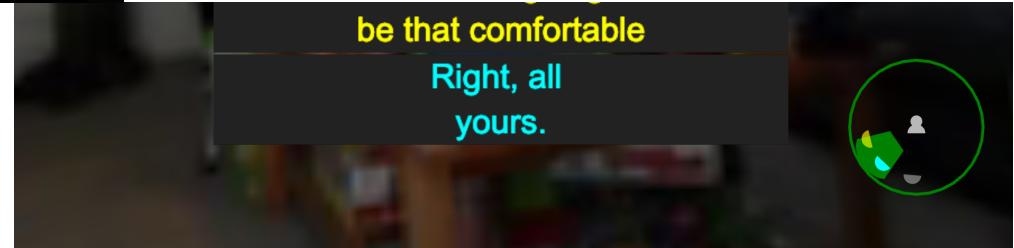
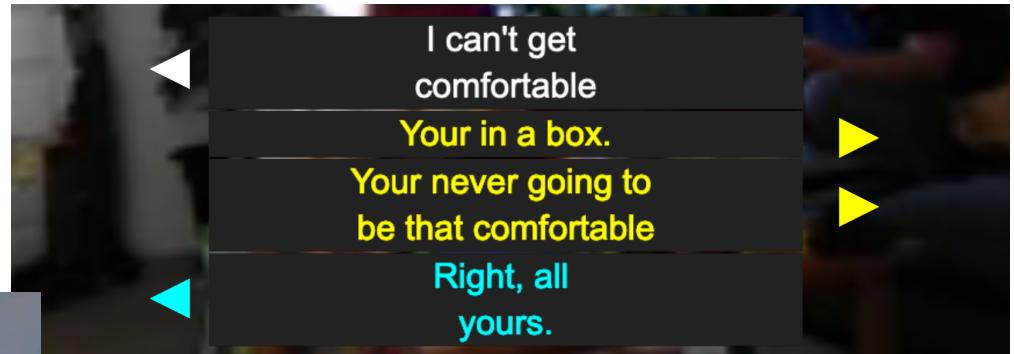
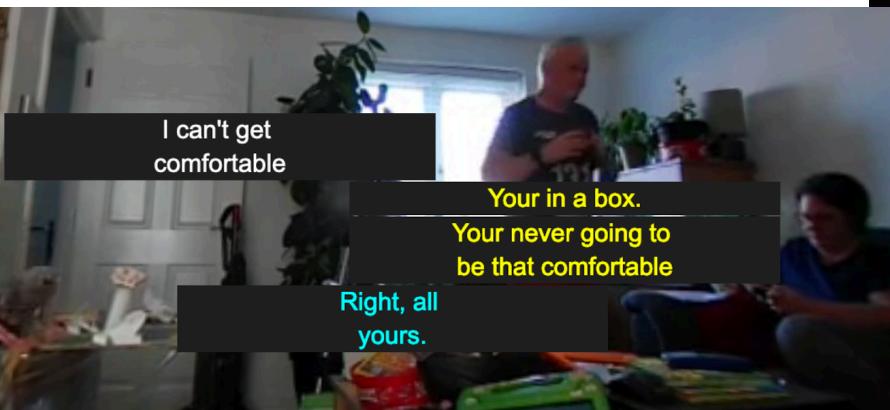
Hier schlägt das  
Herz der Abendschau,  
im zentralen  
Redaktionsbüro.

Hier werden die Reporter  
ausgeschickt und festgelegt,  
welche Themen es in  
die Sendung schaffen.

# Responsive (Customization)



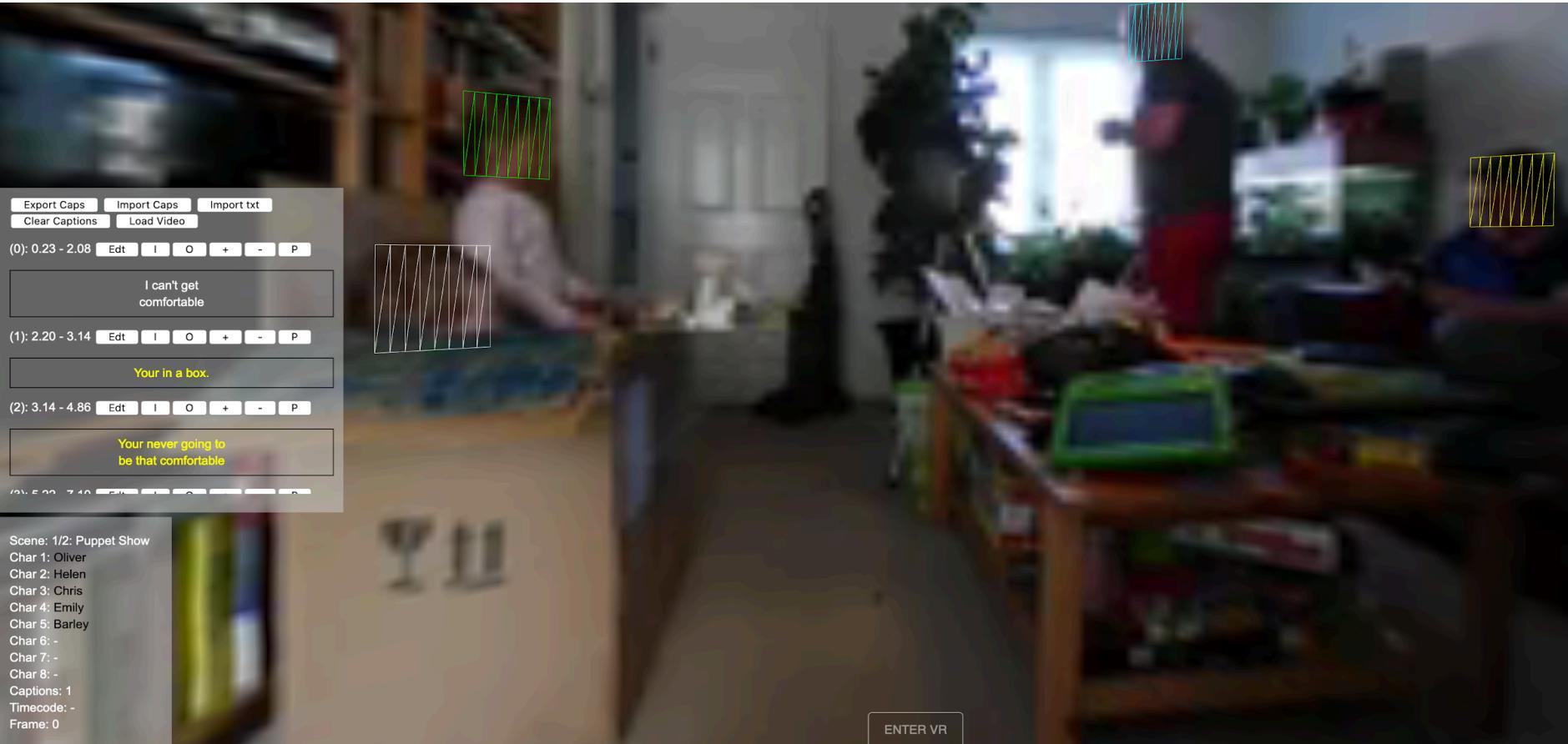
# Extended Guides



# Custom Render (CSS)

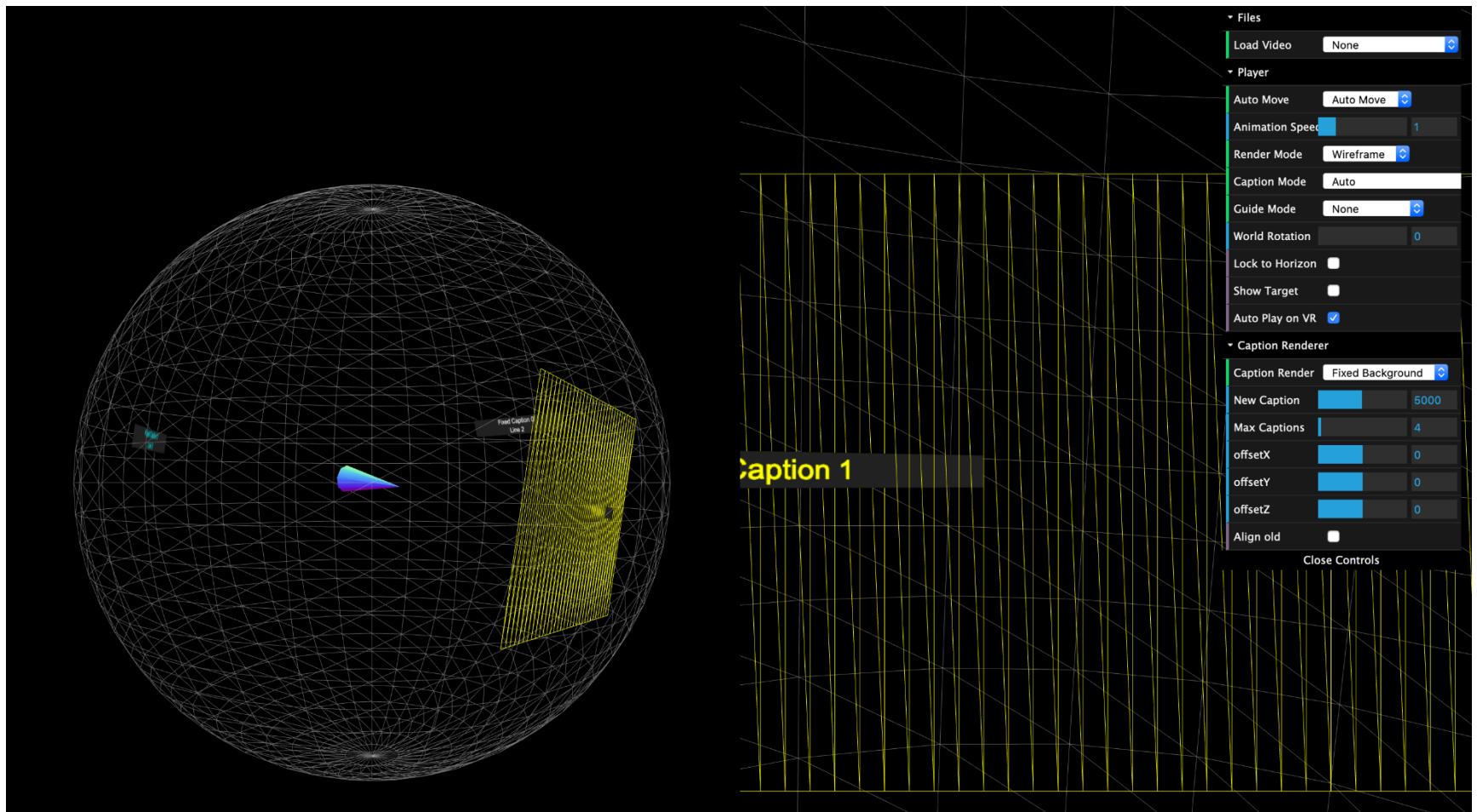


# Tracking and editor



# Try Version 3

- [https://www.chxr.org/immersive\\_subs2\\_3](https://www.chxr.org/immersive_subs2_3)



# References

- Hughes, CJ and Montagud, M 2020, 'Accessibility in 360º video players' , *Multimedia Tools and Applications* .
- Hughes, CJ , Zapata, MB, Johnston, M and Orero, P 2020, Immersive captioning : developing a framework for evaluating user needs , in: *IEEE AIVR 2020 : 3rd International Conference on Artificial Intelligence & Virtual Reality 2020*, 14th-18th December 2020, Online. (In Press)

# Future Work

# Eye Tracking Study



# Conclusions

- User tests yield limited results unless you can put a working product in front of the user
  - If you offer a paper prototype it is possible to cause confusion
  - Often leads to users saying they prefer what they already have.
- In this area many technologies have a learning step
  - You cannot ask a user to evaluate a prototype while they are learning
- The content is important
  - If the users have no interest in what is being presented it will not be a fair test