

# 8th Live Subtitling and Accessibility Symposium Barcelona, 19 April 2023

## Live subtitling: The possibilities that consumers don't yet realize they want



University of  
**Salford**  
MANCHESTER

**Dr Chris Hughes**

Director of Computer Science and Engineering  
University of Salford, UK



Universidade de Vigo

**SWISS TXT**



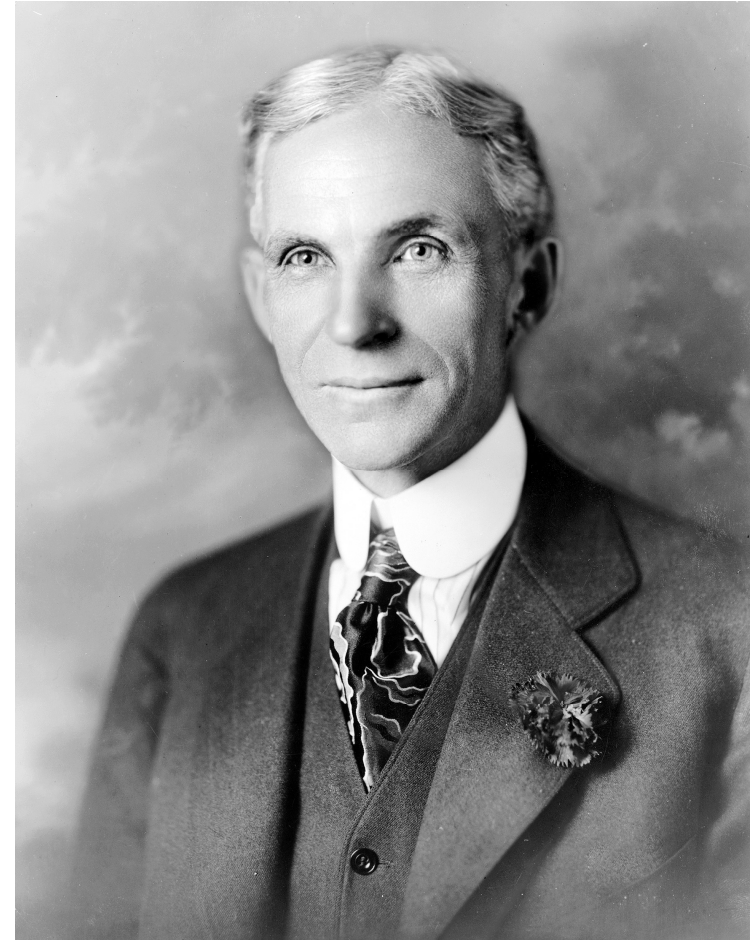
TRANS  
MEDIA  
CATALONIA

---

**“If I had asked people what they wanted,  
they would have said faster horses.”**

- Henry Ford

*(Who never actually said it..)*



# Summary

---

1. Consumers don't know what they want
2. Making a case for prototyping
3. Understanding Consumers
4. The future

# Disclaimer

---

- **Disclaimer 1:** I am a Technologist who has been involved in user testing, not a user test or accessibility expert!
- ***Disclaimer 2:*** *Technologists also very occasionally get things wrong*



Part 1:

Consumers don't know what they want

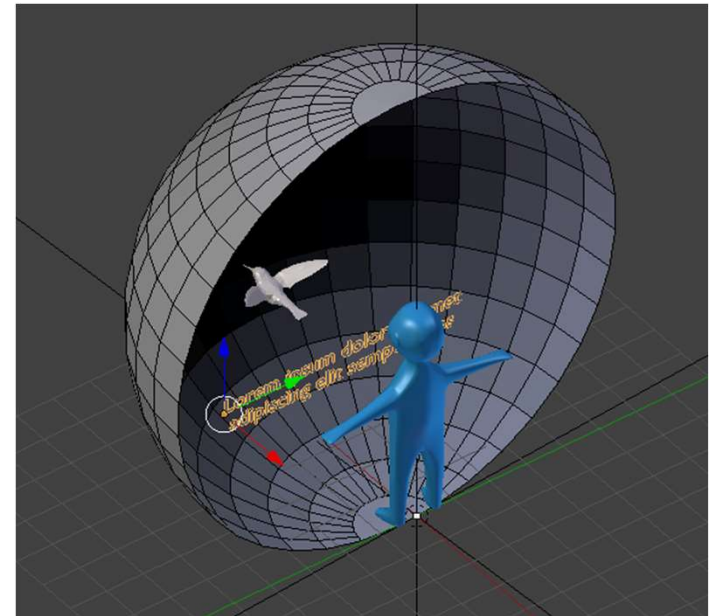
# Case study 1: ImAc (Immersive Accessibility)

**EU H2020:** exploring how accessibility services can be efficiently integrated with immersive media

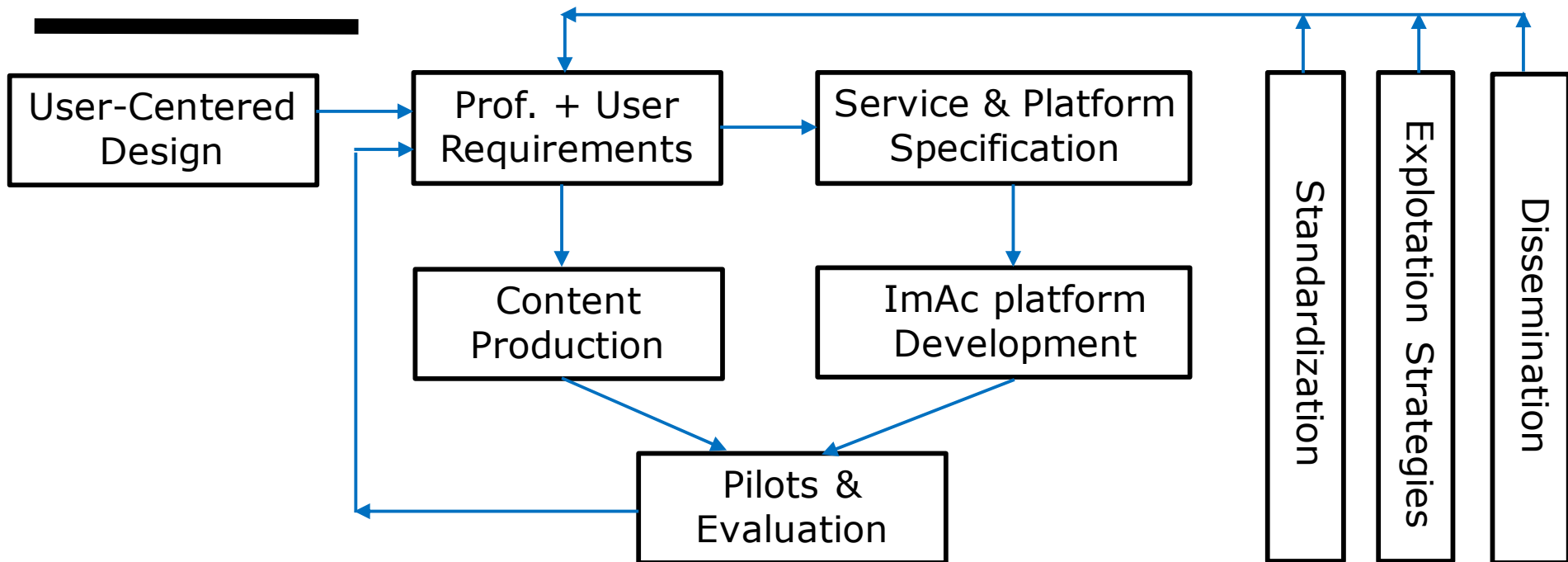
- **Access Services:** subtitles, audio description, sign language
- **Immersive Media:** omnidirectional video (i.e. 360°) and audio

## **Basis:**

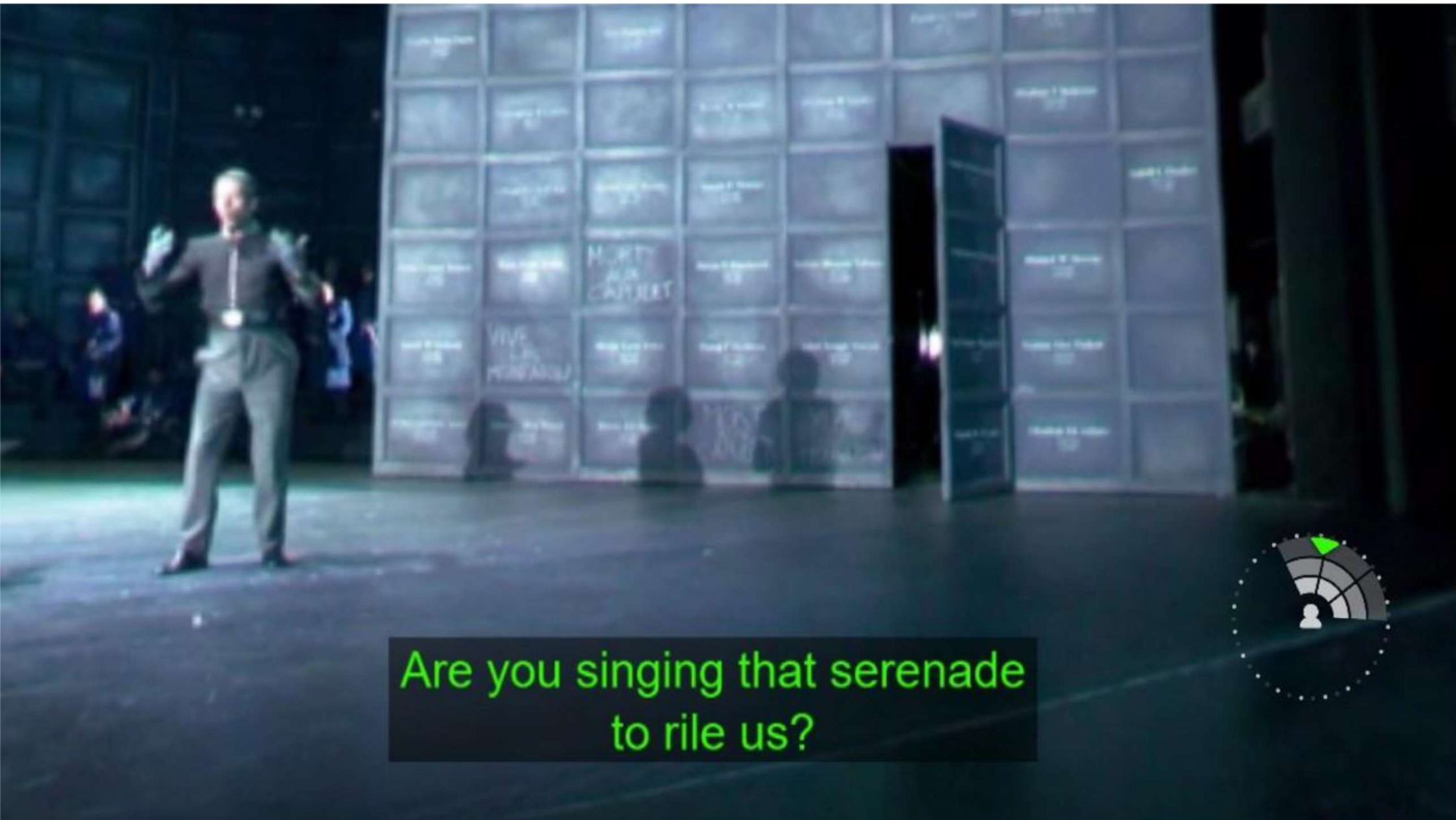
- Accessibility is a must for **e-inclusion**
- Accessibility must **not** be considered as an **afterthought**, but as a key aspect in the specification and deployment of services
- Keep **compatibility** with current standard technologies / formats
- **User-Centric Methodology**



# Case study 1: ImAc (Immersive Accessibility)







Are you singing that serenade  
to rile us?



# Case study 1: Lessons Learned

---

- 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

# Case study 2: W3C Immersive Captions CG

---

- 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 is 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 is challenging
Appear in front, then fixed in scene	Always appear in view before user moves their head don't distract after appearing	May be out of view when looking up/down Fixing in only one direction seems more distracting than follow head
Fixed, position in scene	Clear, consistent position in scene Not needed in other scenes	May be out of view when looking up/down Fixing in only one direction seems more distracting than follow head
Head-locked	Always in view Simple to implement Understandable	May be out of view when looking up/down Fixing in only one direction seems more distracting than follow head May be distracting, feel "stuck to your face" Fixing in only one direction seems more distracting than follow head May cause discomfort 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

**WTF does this all mean????**

Part 2:  
Making a case for prototyping

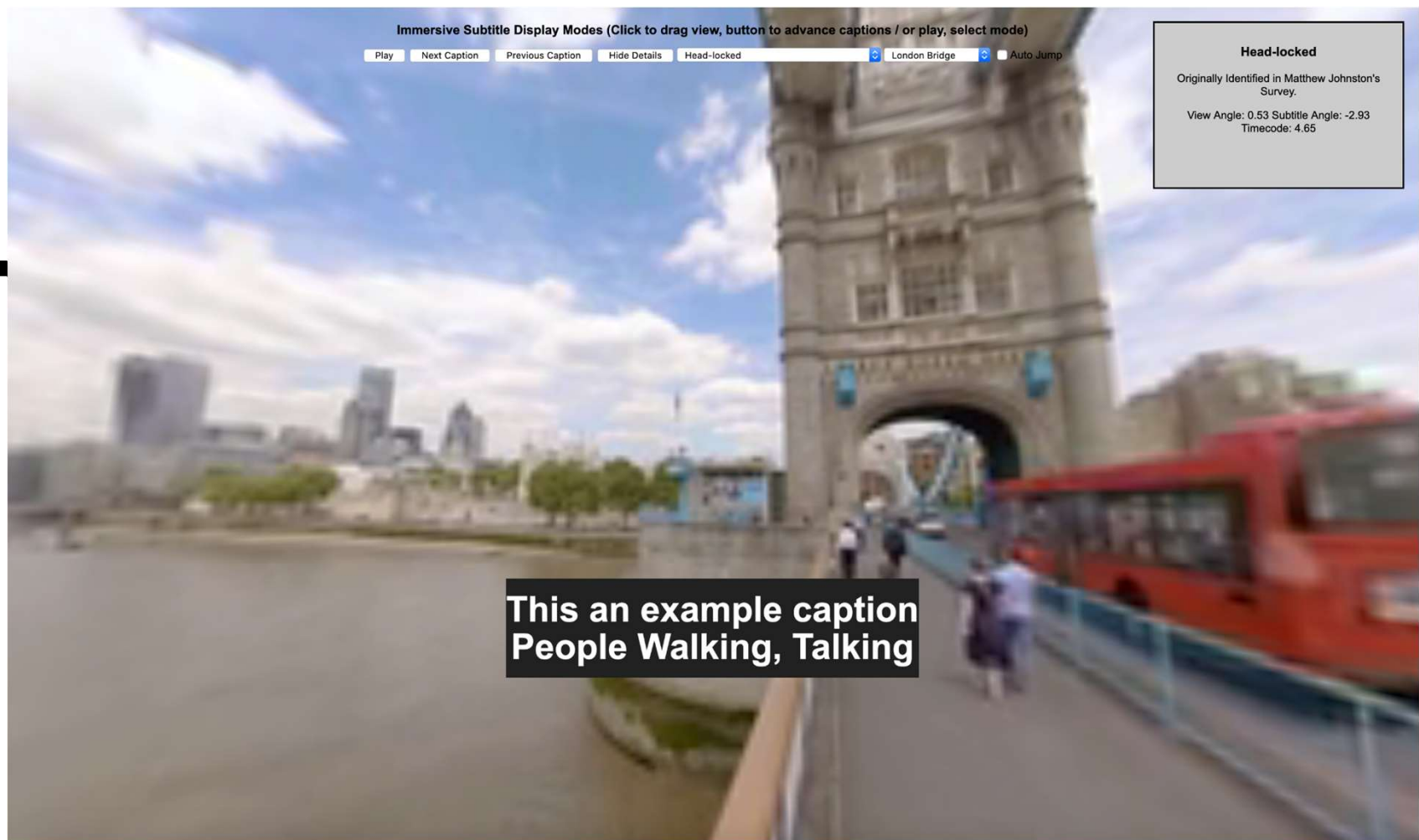
# Prototyping (or hacking)

---

- **What do we think we want?**
  - **How do you know without trying?**
- 

# Hack 1

---



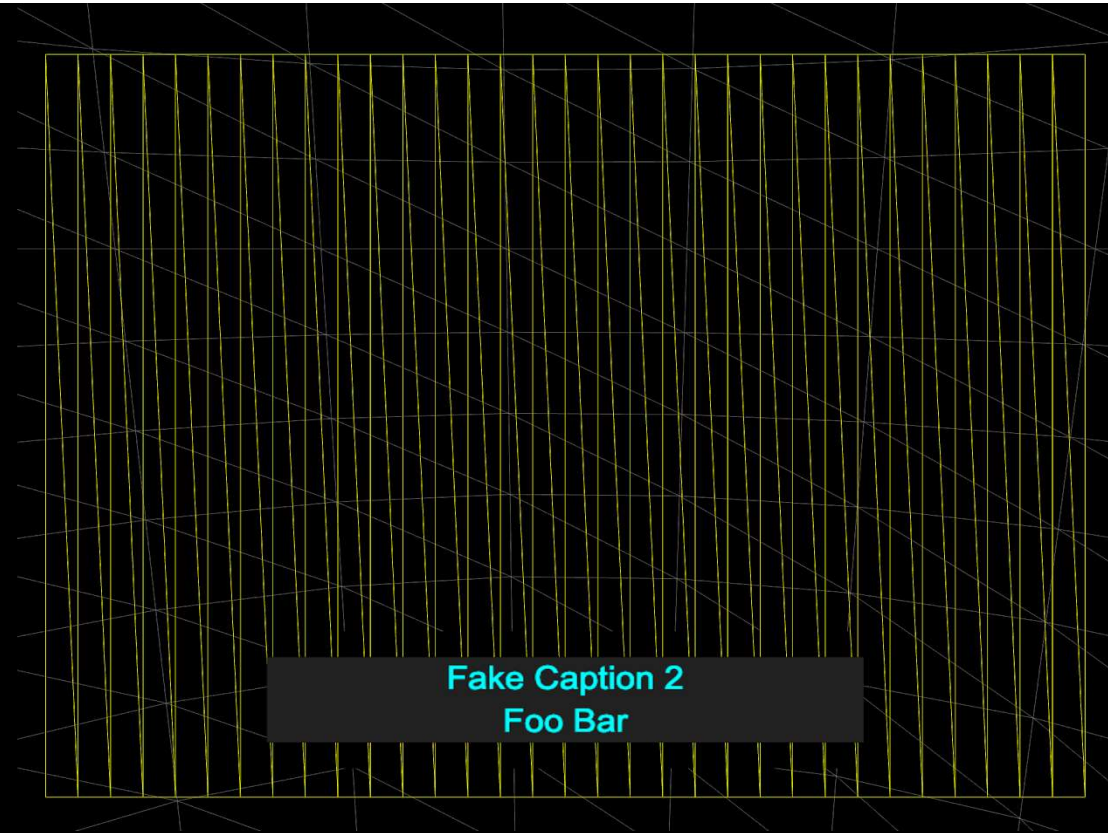
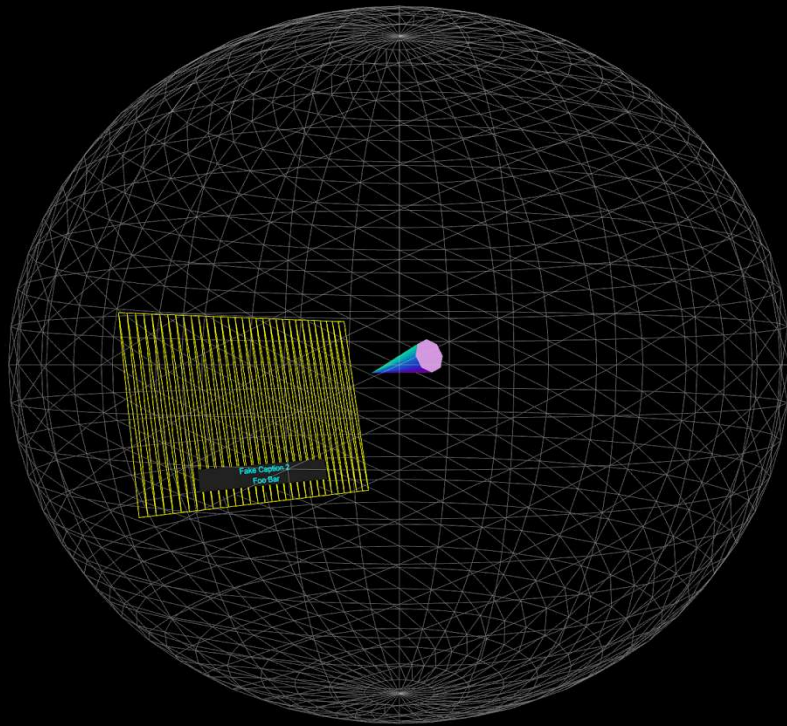
# Hack 1

The screenshot displays the 'Immersive Subtitle Display Modes' interface. At the top, it says 'Immersive Subtitle Display Modes (Click to drag view, button to advance captions / or play, select mode)'. Below this are several buttons: 'Play', 'Next Caption', 'Previous Caption', 'Hide Details', 'Head-locked', 'London Bridge', and 'Auto Jump'. A dropdown menu is open under 'Head-locked', listing various modes: 'Fixed in Scene (Locked Vertical?)', 'Fixed in scene, repeated evenly spaced', 'Appear in front, then fixed in scene', 'Fixed, position in scene', 'Head-locked' (which is selected with a checkmark), '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', and 'ImAc Radar'. To the right of the interface, a box titled 'Head-locked' provides details: 'Originally Identified in Matthew Johnston's Survey.', 'View Angle: 0.53 Subtitle Angle: -2.93', and 'Timecode: 4.65'. The background of the interface shows a blurred video scene with a caption that reads 'This an example caption People Walking, Talking'.

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



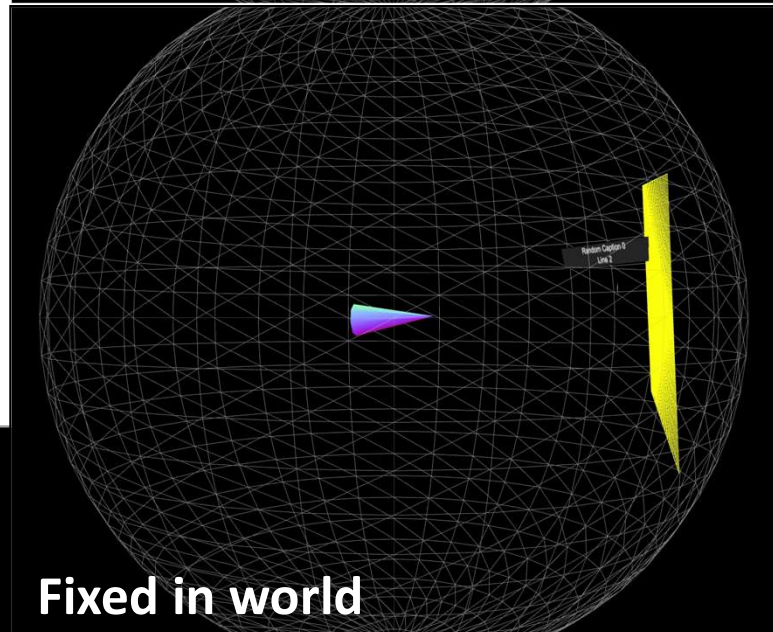
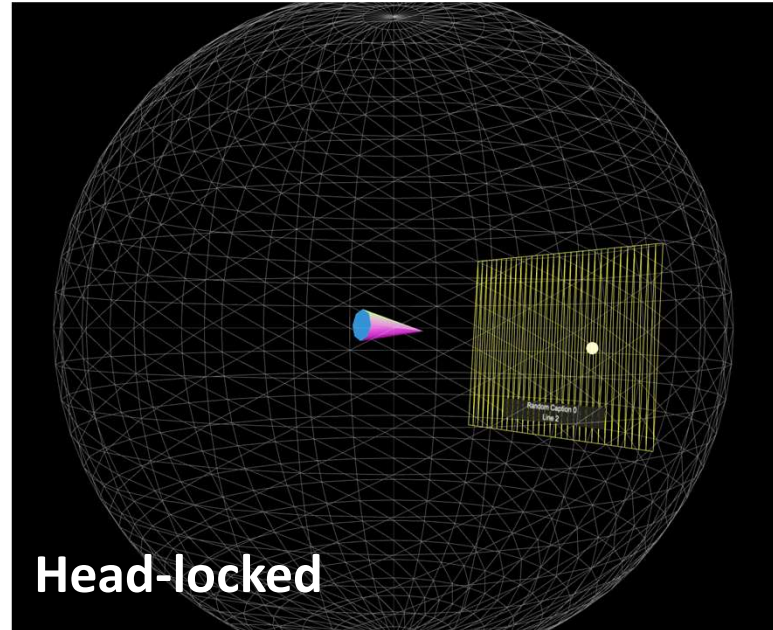
# Hack 2



# Hack 2

---

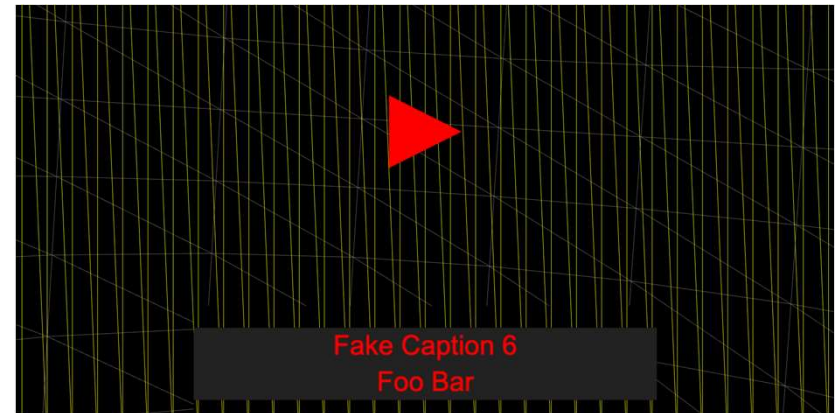
- Fixed in Scene, Locked Vertical
- Fixed in scene, repeated evenly
- Appear in front, then fixed in scene
- Fixed, position in scene
- Head-locked
- Head-locked on horizontal axis only
- Head-locked on vertical axis only
- Head-locked with lag, animate into view
- Head-locked with lag, jump into view



# Hack 2 - Guide Modes

---

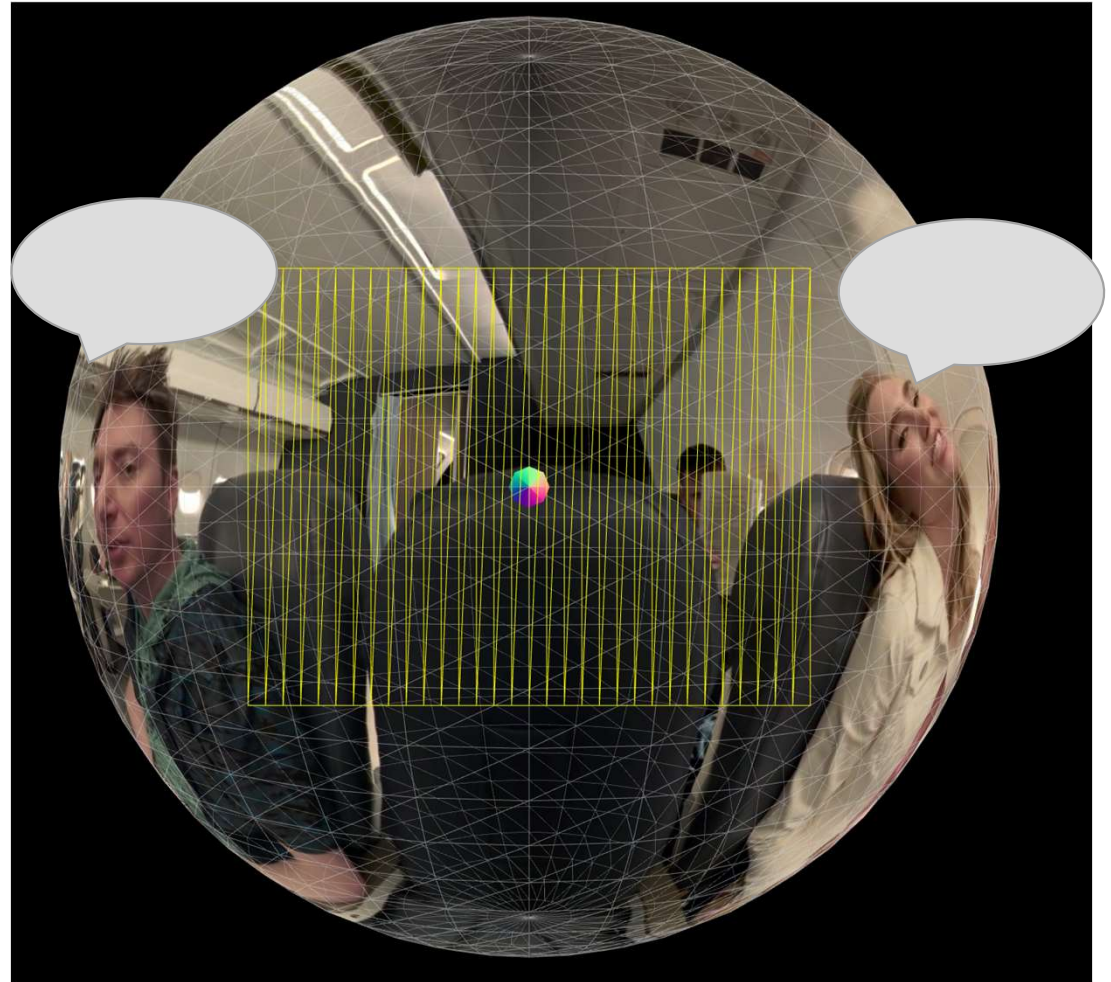
- Guide Modes were replicated





# Hack 2 – Perfect?

---

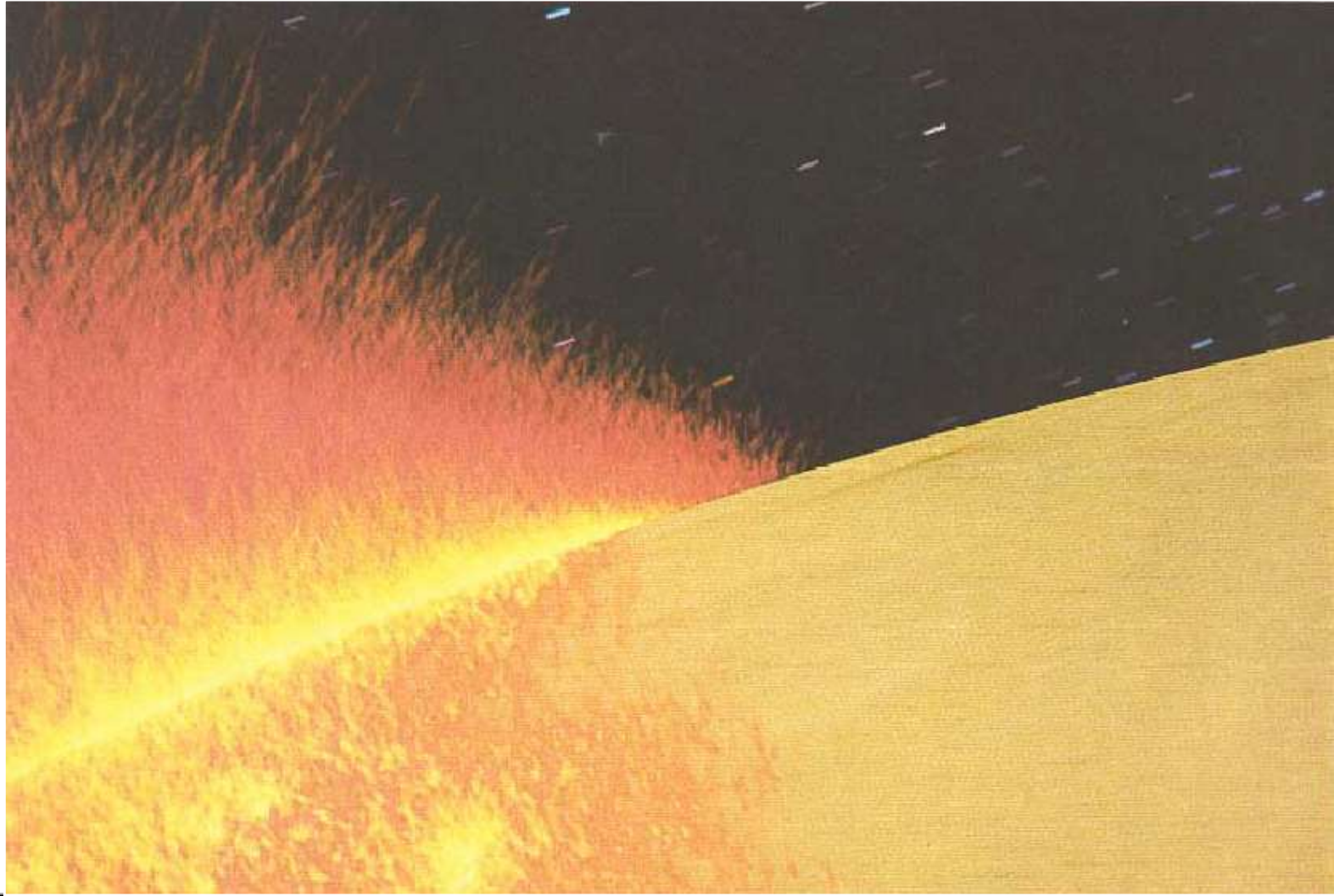


# Hack 3

---

## Particle Systems

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

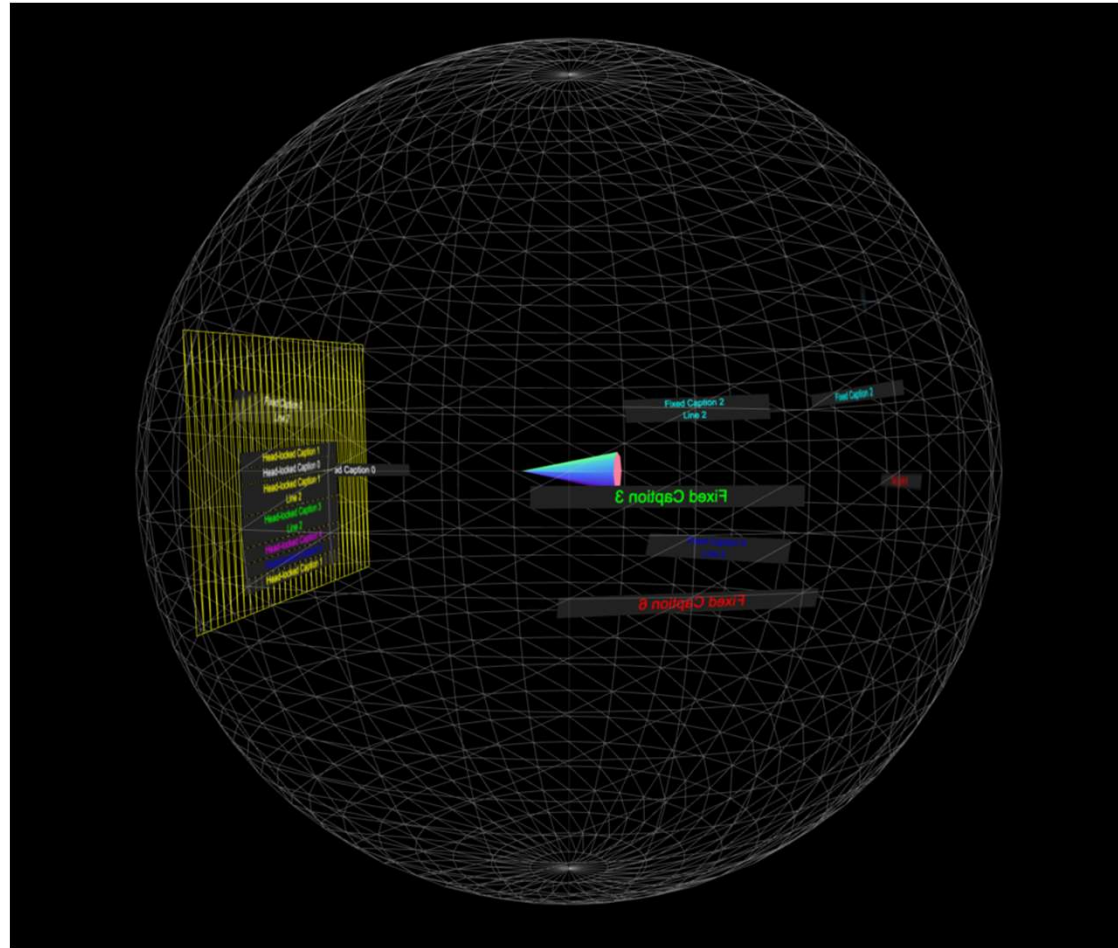


# Hack 3

---

## Captions Object based

- Caption Emitter
- Caption Manager





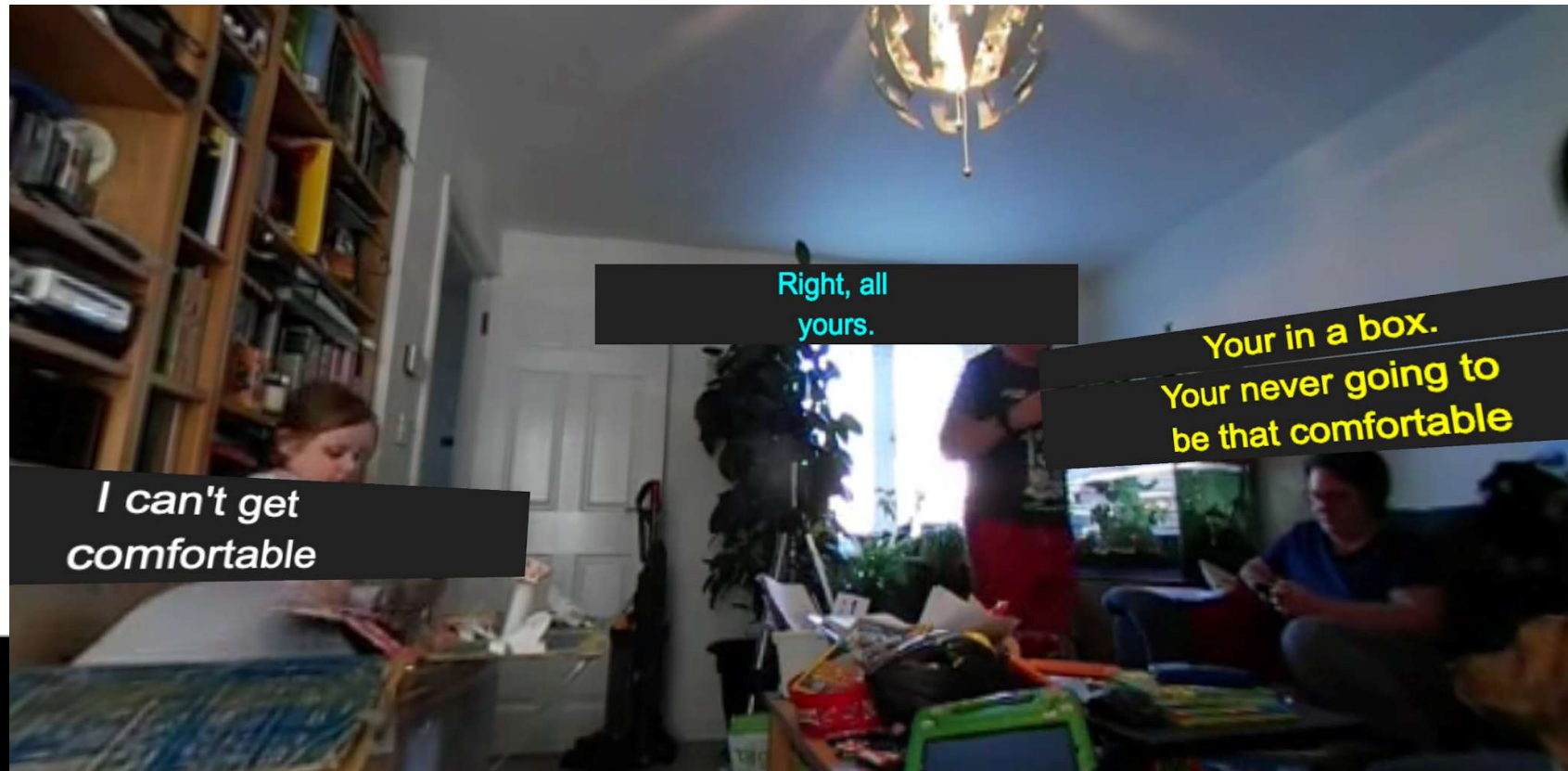
Home | Back | Play | Pause | Stop | Seek Forward  
Next Caption | Previous Caption | Clear Captions | Show Options



VR NOT SUPPORTED

Playback Rate: 1.0x  
Auto Move: Auto Move  
Animation Speed: 1  
Render Mode: Fullscreen VR  
Caption Mode: Auto  
Audio Mode: None  
World Rotation: 0  
Lock to Holo:   
Show Target:   
Auto Play on VR:   
+ Responsive Captions  
Enabled:   
Size: 5  
+ Timescale  
Enabled:   
Format: Text  
Position: Middle  
Style: BlackWhite2D  
+ Caption Renderer  
Caption Render: Text Displacement  
Caption Justify: Center  
Curve: 0  
BackOffsetX: 0  
BackOffsetY: 0.25  
BackOffsetZ: 0  
HatchedOffsetX: 0  
HatchedOffsetY: 0  
HatchedOffsetZ: 0  
Character Name:   
Only on Change:   
Use Icon:   
Caption Threshold: Old Mac  
Removal Delay: 2000  
Max Captions: 1  
Align old captions:   
Transcript:   
Maximum: 7  
Curve: 0  
Transcript Box:   
+ Tracked Location  
Target set from: From caption box  
Show Tracking:   
Call Mode:   
+ Caption Editor  
Caption Editor:   
Scene Info:   
Close Dialog

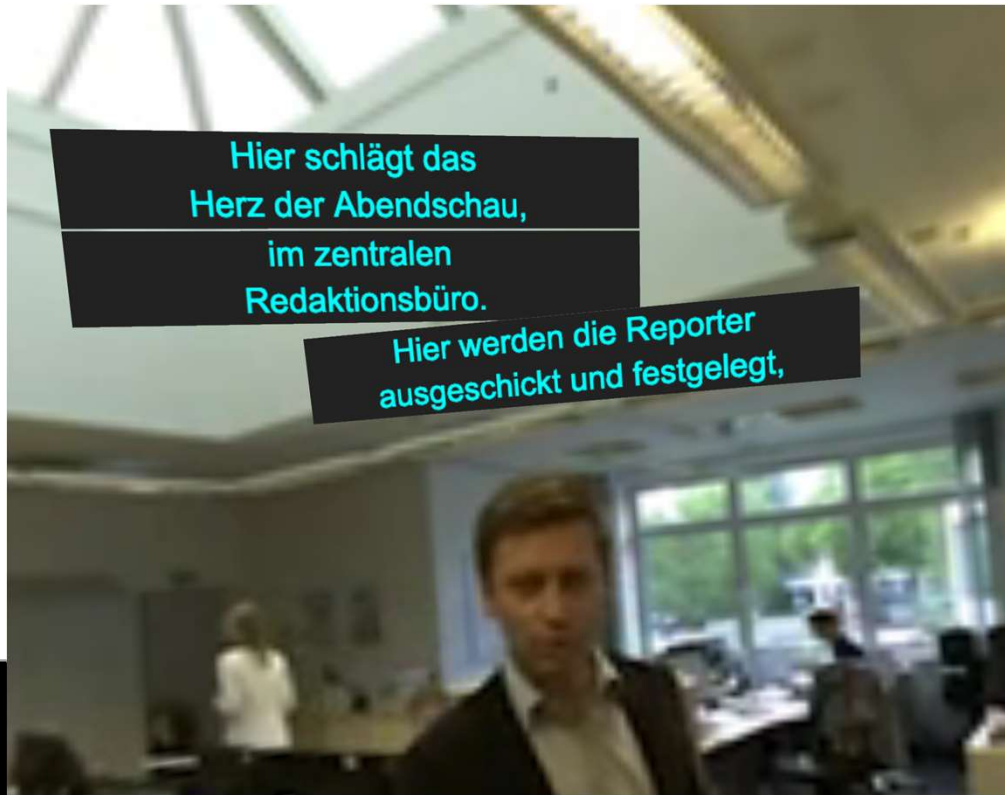
# Captions fixed to speaker





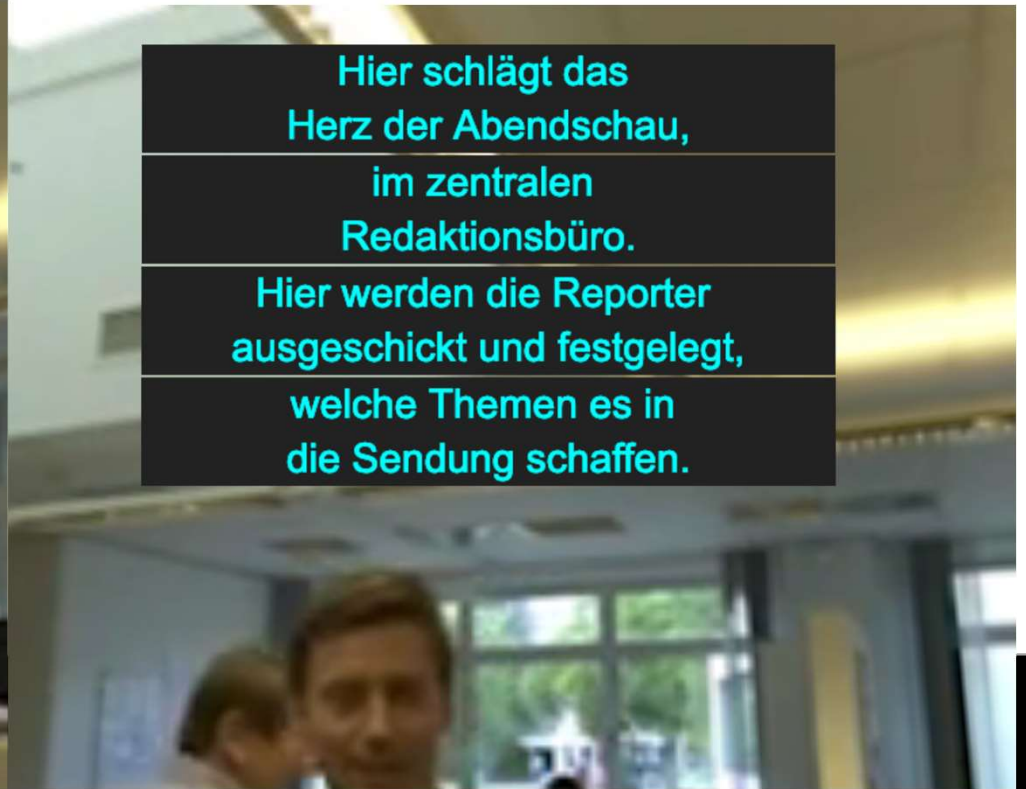
# Stacking (Collisions)

---



Hier schlägt das Herz der Abendschau, im zentralen Redaktionsbüro.

Hier werden die Reporter ausgeschickt und festgelegt,

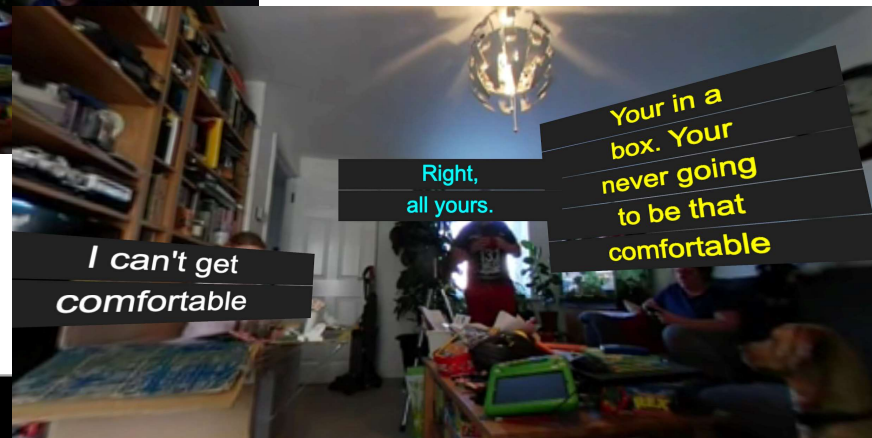


Hier schlägt das Herz der Abendschau, im zentralen Redaktionsbüro.

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

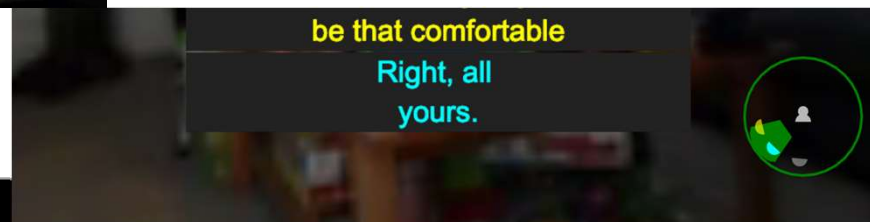
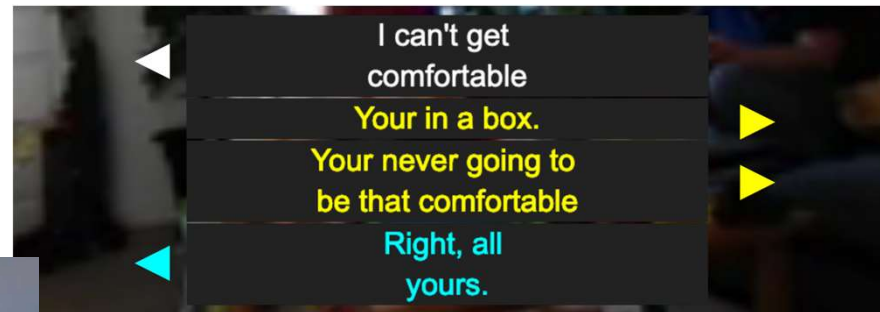
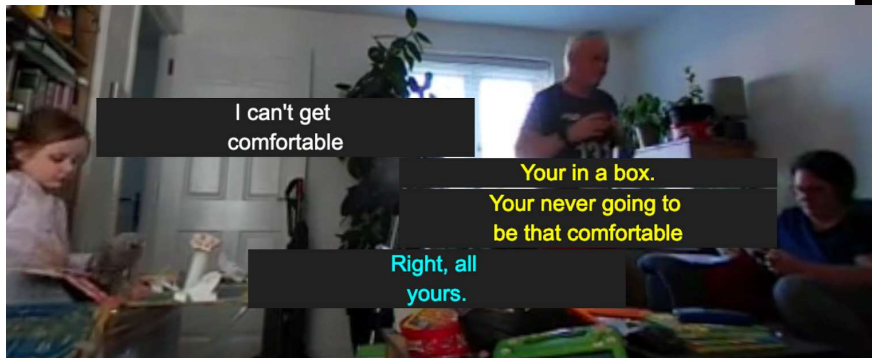
# Responsive (Customization)

---



# Extended Guides

---



# Custom Render (CSS)

---



# Tracking



# New Observation

---

- Users don't know what they do:
  - If you ask them if they read a subtitle or saw an object they often don't know if they did or not.
  - Possibly these tasks are done subconsciously.
  - Possibly the user doesn't remember.
- Eye tracking doesn't lie...

Part 3:  
Understanding Consumer Behaviour

Playback



# HTC Vive Eye Pro

---



# HTC Vive Eye Pro

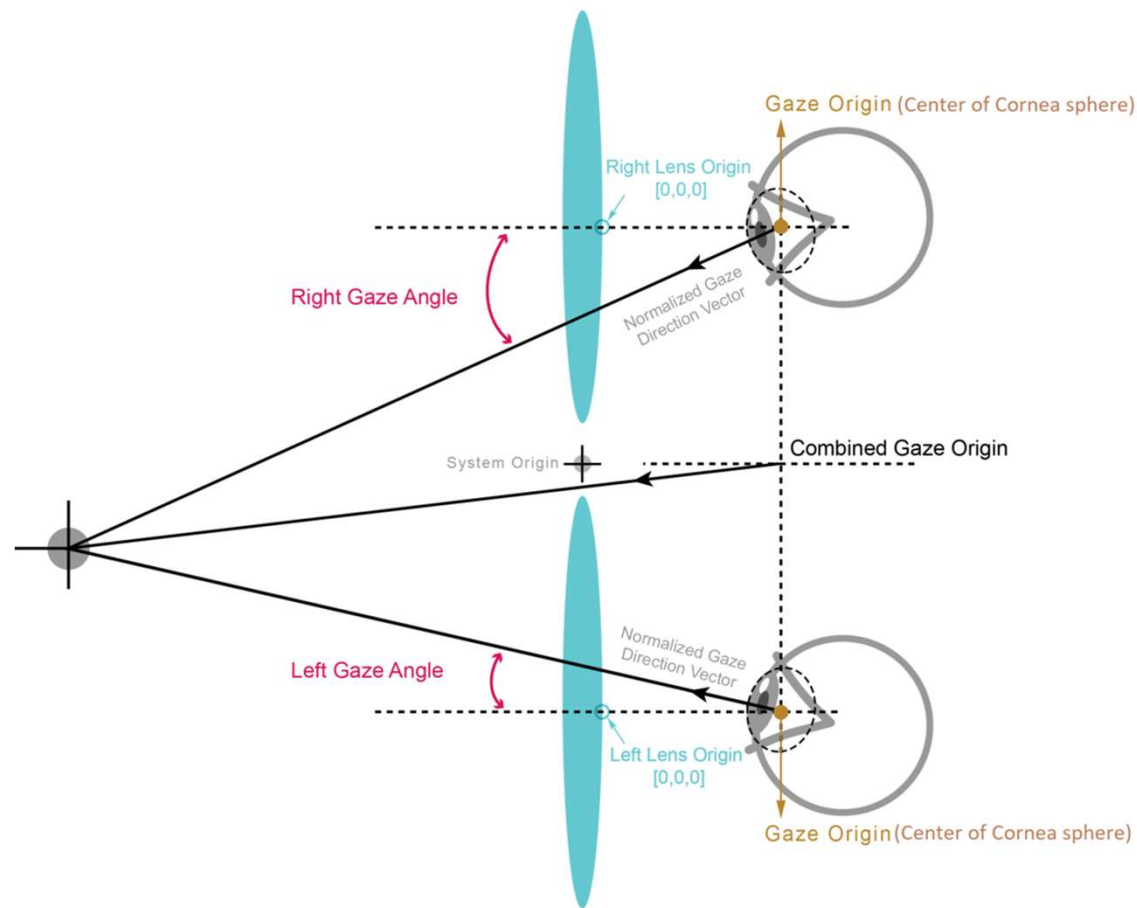
---



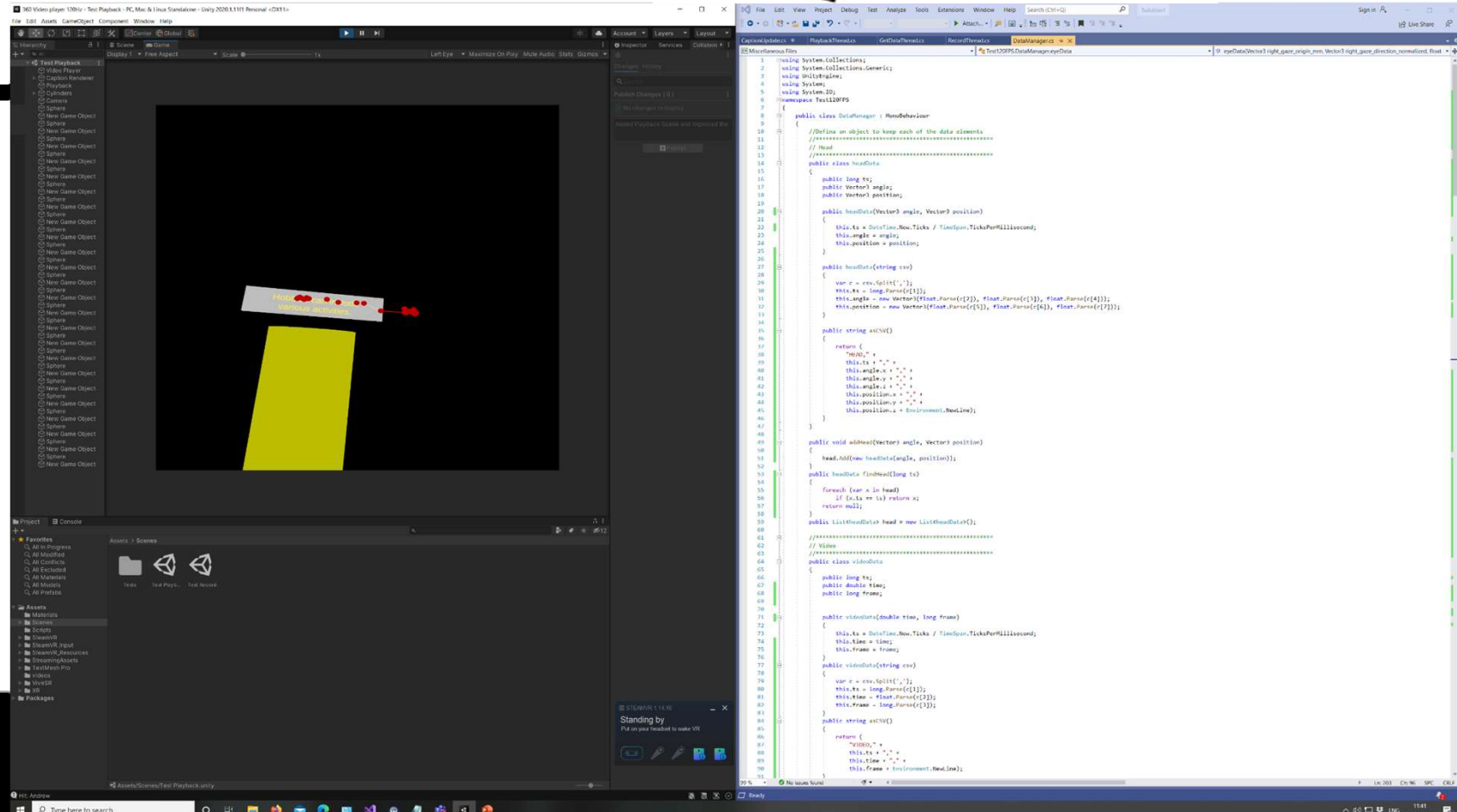
- 120Hz Eye tracking

# HTC Vive Eye Pro

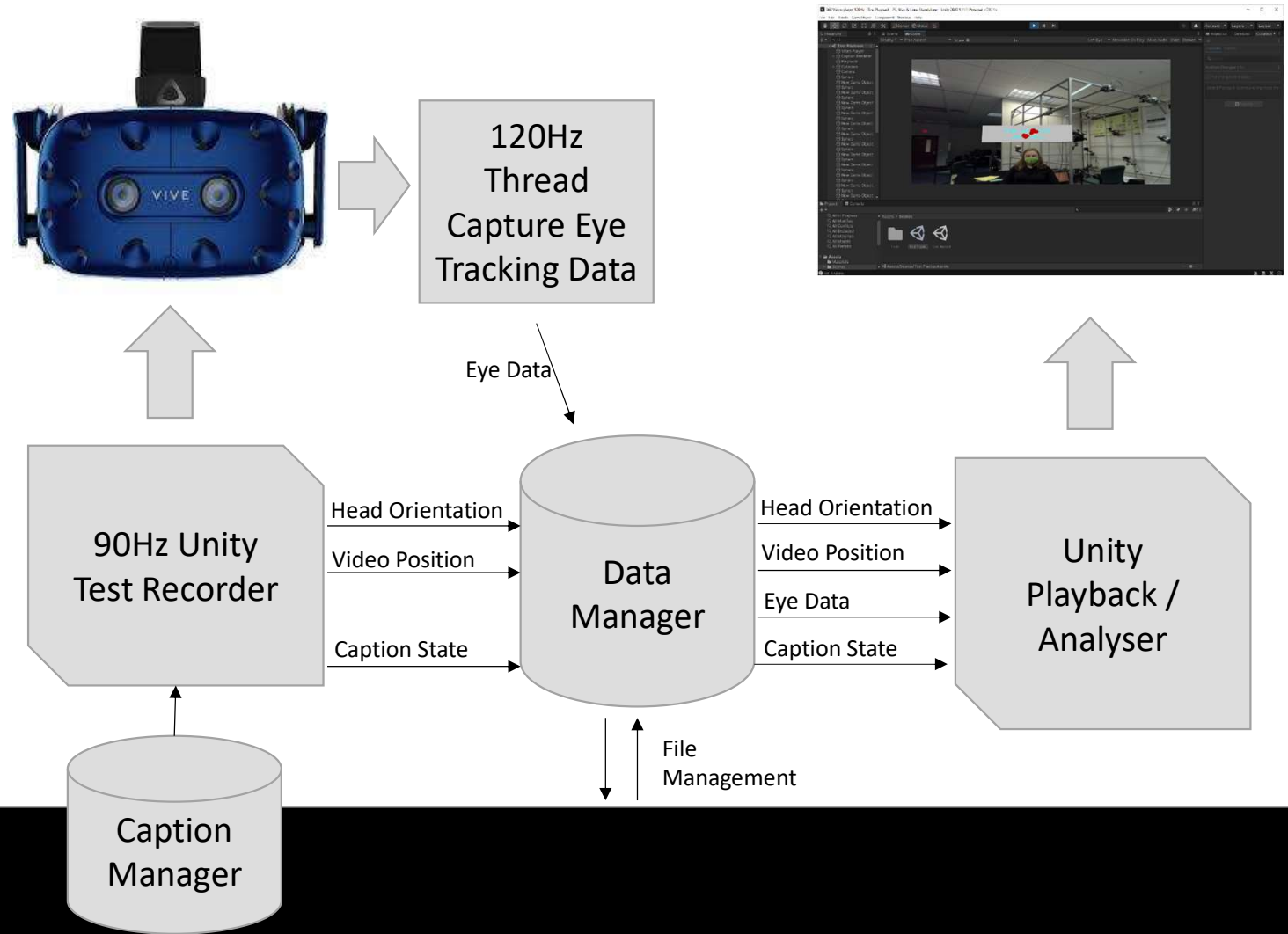
---



# Unity



# Architecture

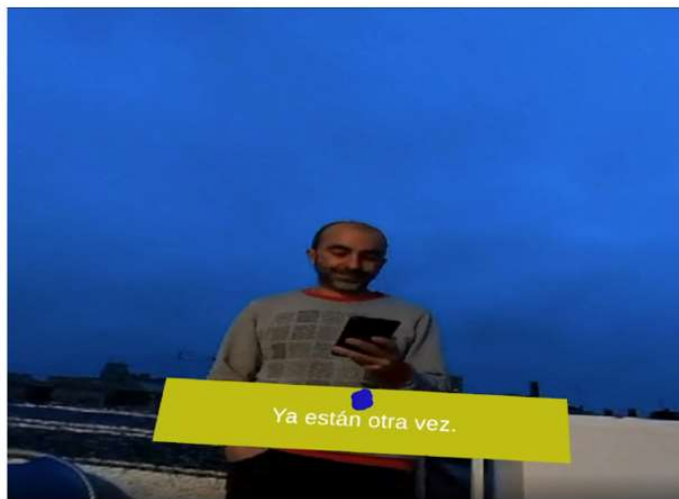




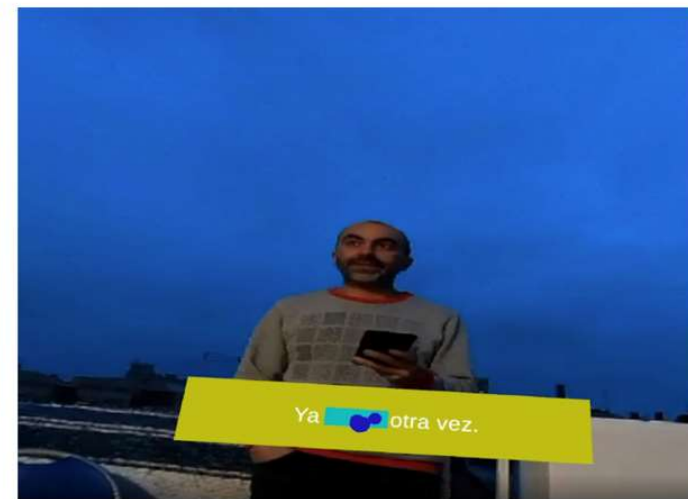
(a) Actor body AOI



(b) Subtitle quad AOI



(c) Subtitle word AOI



(a) saccade flight



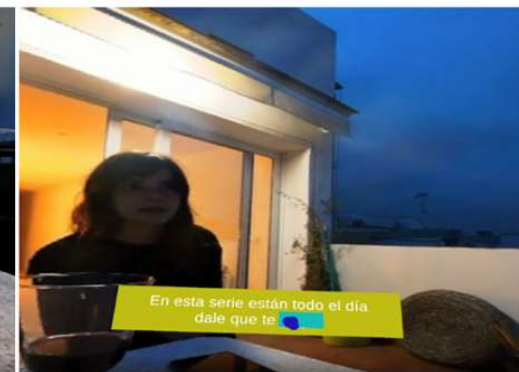
(b) saccade landing below text



(c) saccade landing on text



(d) fixation on word





Pause on Caption



Pause on Read

# Ambient Audio

1 00:00:06.1  
1.19  
00:00:07.0  
37 00:00:03  
1.78 Ya están otra vez.

2 00:00:09.890  
1.97 ¡Ora lo dije!

44 00:00:10.890  
3.15 ¡Esta será esta! ¿Arate!

00:00:14.160  
2.05 No, no, me refiero a los vecinos.

3 00:00:15.330  
2.23 ¡Quiero ir a casa!

4 00:00:17.050  
0.99 ¡Quiero ir a casa!

38 00:00:20.170  
0.99 ¡Quiero ir a casa!

5 00:00:22.420  
2.78 ¡Ya está bien!

00:00:24.670  
1.53 ¿Quieres llamar a la policía?

6 00:00:24.790  
2.00 Deja que le marque.

39 00:00:25.320  
1.16 ¡Vale!

00:00:30.870  
1.00 ¿Sí? Hola.

41 00:00:32.520  
1.80 Un momento...

7 00:00:36.360  
2.90 La paso a mi compañera de piso.

42 00:00:40.160  
2.99 ¡Hola, ¿cómo estás?

8 00:00:44.090  
2.73 ¡Hola, ¿cómo estás?

29 00:00:49.010  
3.80 ¿Qué ha pasado? ¿Qué haces?

00:00:55.890  
3.80 ¿Qué ha pasado? ¿Qué haces?

10 00:00:55.890  
1.72 ¡Quiero ir a casa!

28 00:00:58.490  
1.72 ¡Quiero ir a casa!

00:01:01.210

Select  
All None First  
Last Next Previous  
Next Previous  
1 2 3 4 5 6 7 8  
Set Character  
1 2 3 4 5 6 7 8  
Captions  
Clear All Delete Trim  
Divide Combine Split  
Orientation  
Clear  
File  
Commit  
Browse... No file selected.  
History  
Undo Redo

Play Scene Play Pause Stop << >> | Save Amb Load Amb Auto Pause

Joder!

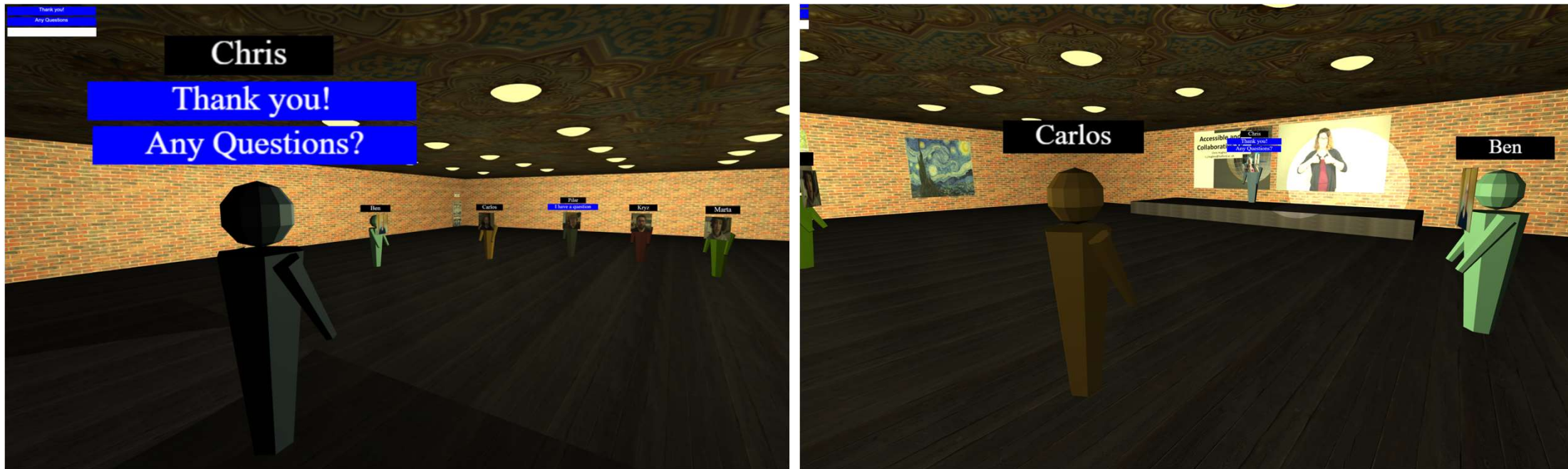
Ya están otra vez.

¡Ora lo dije!

Scene 1

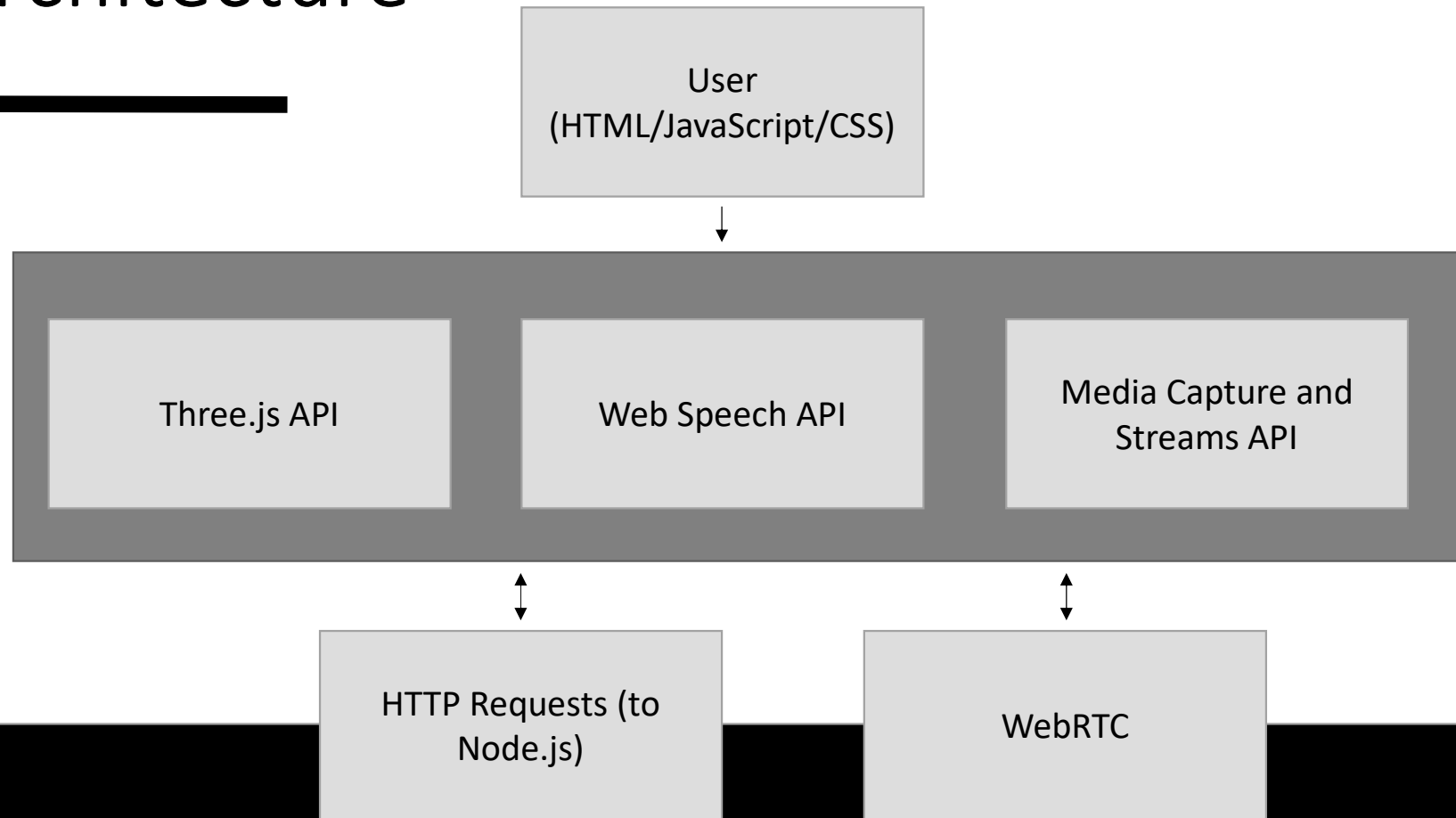
Part 4:  
Future uses of Live captions (XR)

# Live Subtitles in Collaborative XR



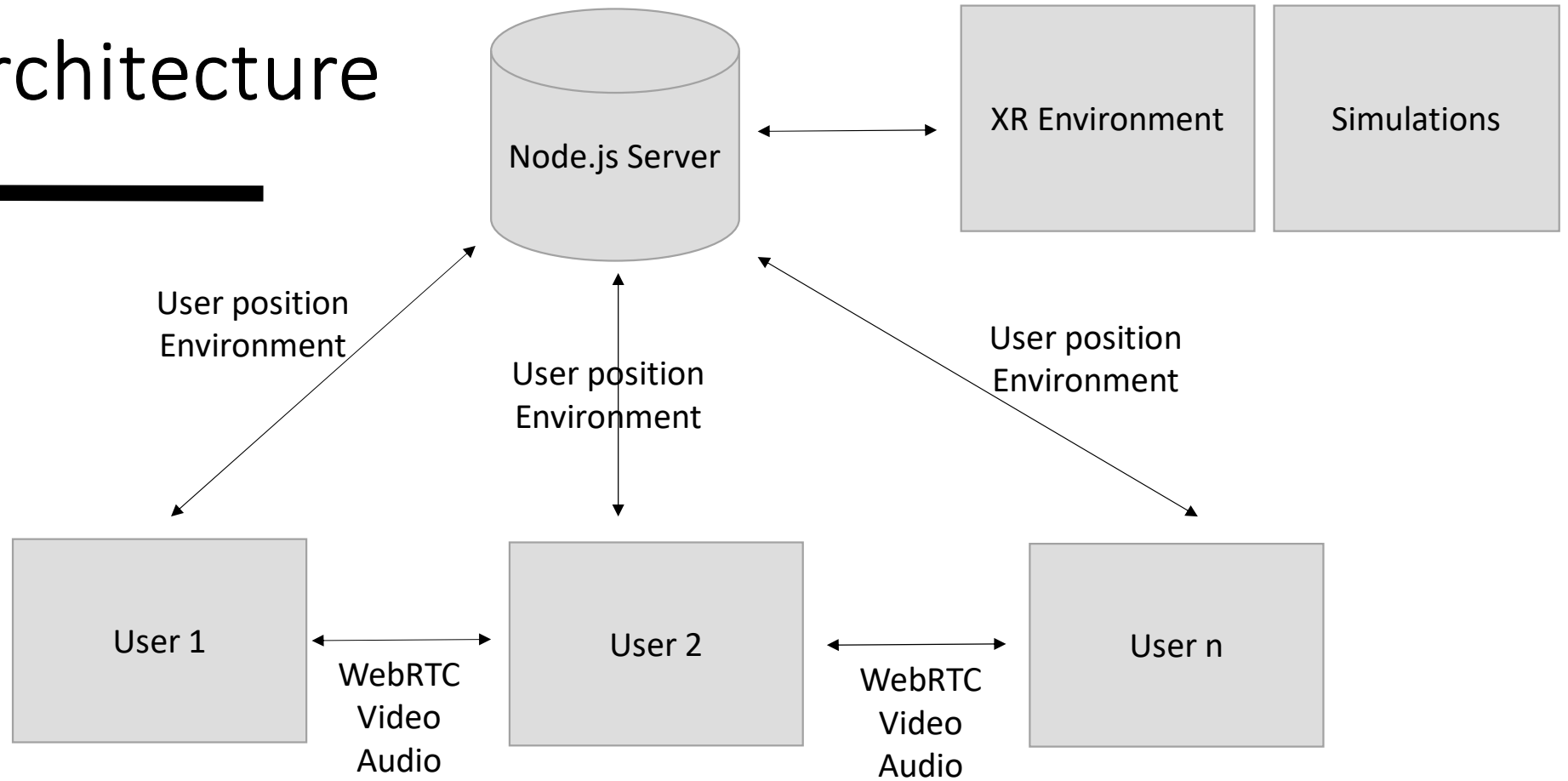
# Architecture

---



# Architecture

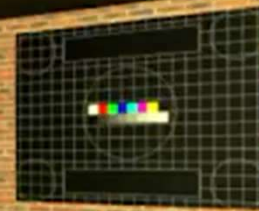
---



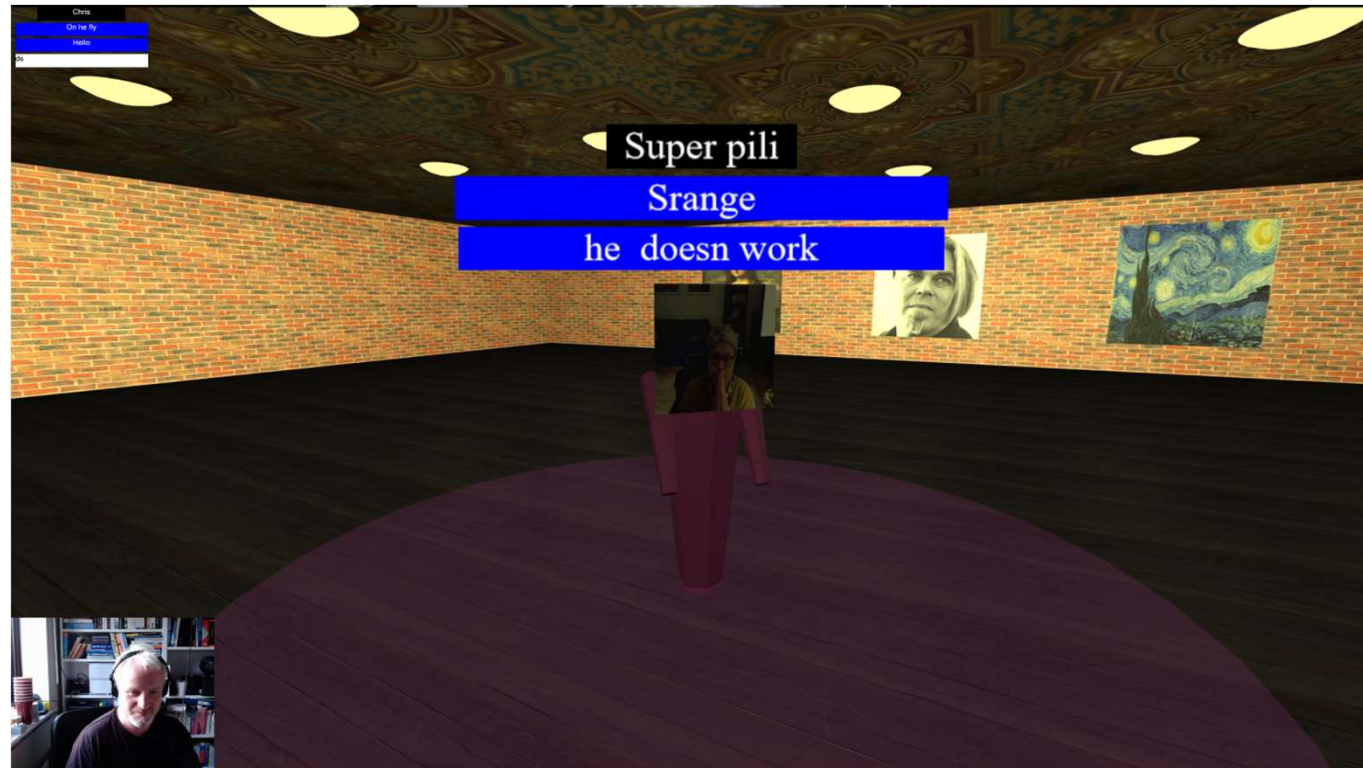


Collaborative  
XR

Chris



# Technologists *sometimes* make mistakes





# Conclusion

---

- User tests yield limited results unless you can put a working product in front of the user;
- Many technologies have a learning step;
- The content is important;
- Users often don't even know what they do;
- Some cool ideas come from experiments;
- *although.. Technologists often get it wrong and need users to tell them!*

# 8th Live Subtitling and Accessibility Symposium Barcelona, 19 April 2023

## Live subtitling: The possibilities that consumers don't yet realize they want



University of  
**Salford**  
MANCHESTER

**Dr Chris Hughes**

Director of Computer Science and Engineering  
University of Salford, UK



Universidade de Vigo

**SWISS TXT**



TRANS  
MEDIA  
CATALONIA