# Improving Accessibility for Deaf Audiences in Extended Reality Theatre Productions: A proposed approach to developing Advanced Captioning Solutions

**>> StageTEXT**

Written by:

**Stagetext and Peak Signal**

**PEAK SIGNAL**

# Table of Contents

SEPTEMBER 2024

# Introduction

More and more artists and theatre makers are experimenting with presenting work in extended reality[1]. These theatre productions use technology to transform the way audiences experience performance, creating a deeply engaging and interactive environment.

Stagetext is a deaf-led charity that advocates for and provides accessible captioning[2] for live and recorded theatre. We are passionate about ensuring all types of theatre performances are accessible to deaf, deafened and hard of hearing audience members.

We see the evolution of extended reality performance as an opportunity for access to be considered and embedded at the earliest stage. With support from Innovate's Audience of the Future 2 fund, we embarked on research with our technology consultants, Peak Signal, to explore the challenges and opportunities presented by captioning in extended reality.

This whitepaper presents the results of our research with Peak Signal. Stagetext provided the expertise in captioning, and Peak Signal provided the technological expertise on extended reality, and technical insights which, underpinned by the principles of accessibility, form the body of this research paper.

This paper provides insights on the principles for developing captioning processes for extended reality theatre. It provides information on foundational concepts before detailing specific requirements and proposed methodologies for captioning in extended reality, derived from our research. Ultimately this research led to Peak Signal identifying an approach to providing repeatable captioning solutions in extended reality environments.

We share these findings so that artists and theatre makers can build upon our research to create wonderful extended reality work that is accessible to deaf, deafened and hard of hearing people.

[1]  Extended reality (XR) is a term that covers a wide range of technologies, including virtual reality, augmented reality, and mixed reality. Extended reality includes any digitally mediated immersive experience, whether its fully virtual or involves real world interaction

[2]  Captioning is like subtitles, but in theatre. Our captions look similar to subtitles on TV or film, but are displayed on large screens at the front of the theatre to ensure deaf, deafened and hard of hearing audiences can enjoy the show alongside everyone else

# Background and context

Extended reality theatre productions differ from traditional in-person theatre by using technology to break the "fourth wall" and engage the audience directly within the performance space. This interactive approach enhances the sensory experience and deepens audience engagement. This presents unique accessibility considerations for all artists and theatre makers wanting to provide equal access for those who rely on captions to follow a performance. In the UK, this includes 18 million deaf, deafened or hard of hearing adults, plus many more groups who use captioning for literacy, concentration and learning.

Extended reality is a term that covers a wide range of technologies, including virtual reality, augmented reality, and mixed reality. Extended reality includes any digitally mediated immersive experience, whether it's fully virtual or involves real world interaction.

In theatre, extended reality could mean an audience watching or interacting with live or recorded performance through a fully immersive virtual reality headset, through a phone that augments real life performance, or via a wide range of other interactive technologies.

Our objective in this research was to understand how much of the process for creating captions for in-person theatre productions applied in extended reality, to identify gaps, and to look at how to address those gaps.

Stagetext has an established set of key principles about what makes text accessible for scripted in-person theatre. These have been developed through research with deaf, deafened and hard of hearing people over Stagetext's 24 year history.

The key principles for accessible captioning of in-person theatre are:

- Captions should be verbatim – matching speech word for word
- Captions include all the information – like character names and factual descriptions of sounds and music
- Captions should be in time – matching speech as it is spoken
- Captions should be legible – visible and readable to the audience as they take in the performance[3]

It was clear that these principles would apply in extended reality spaces. All of the principles are needed to provide equal access, but in particular the legibility, visibility and readability of captions in virtual spaces where audiences can change their perspective presented an interesting new challenge.

[3]     For more information on these principles, visit

https://www.stagetext.org/wp-content/uploads/2022/11/resources_st_Captioning_Display_Standards-1-1.pdf

When undertaking our research, we were keen not to reinvent the wheel, so there were some key inputs we used:

- The ImAc (Immersive Accessibility) project[4]. ImAc is developing and testing appropriate solutions for both rendering modes and guiding methods for subtitles in 360° videos. This includes considering the impact on user immersion and helping users to understand the story better.
- Existing research on subtitle placement in a virtual environment (including avoiding VR Motion sickness in relation to fixed captions), primarily BBC Research & Development's whitepaper.[5]

Extended reality theatre productions demand innovative solutions for captioning, as the traditional methods of fixed-position captions are not appropriate. In a 360° environment, fixed-position captions can be difficult to follow because they do not move with the audience member's changing perspective. This can lead to situations where captions are out of view or require the viewer to constantly shift their focus, detracting from the immersive experience as well as potentially contributing to VR motion sickness.

# Research approach

Our research followed a structured approach:

**Literature Review**
Examining existing research, including the ImAc project and BBC R&D whitepaper.

**Engagement with theatres**
Examining what categories of productions had been undertaken and if/how accessibility had been provided within extended reality productions.

**Gap Analysis**
Identifying gaps in current theatre captioning processes when applied to extended reality theatre.

**Proposal for how gaps could be closed**
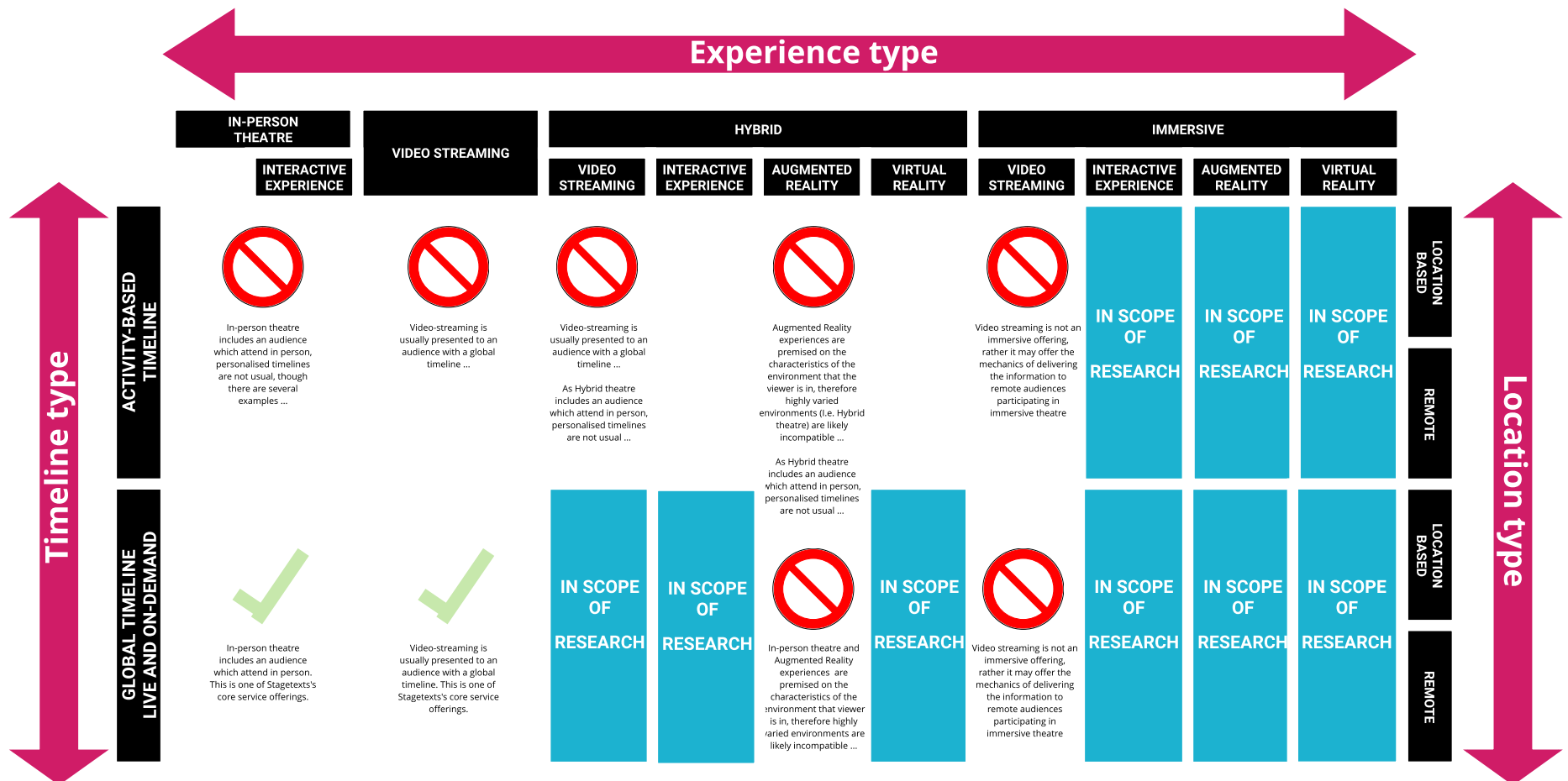Potential solutions for creating and delivering captioned experiences in extended reality environments.

[4]     News – ImAc delivers novel resources for the broadcasting industry to provide adapted content ensuring accessibility in immersive environments (imacproject.eu)

[5]     Exploring Subtitle Behaviour for 360° Video - BBC R&D

# Summary of key findings

The **literature review** and **engagement with theatres** provided a foundation of knowledge on the current state of extended reality captioning and highlighted areas where further development was needed. This included a view on the types of extended reality theatre productions taking place and how audiences were interacting with them.

The below graphic illustrates which type of productions were in scope of extended reality theatre research and therefore which will potentially benefit from advanced captioning solutions to make productions accessible.

**Experience type**

**Timeline type**

**Location type**

| | IN-PERSON THEATRE | VIDEO STREAMING | HYBRID | | | | IMMERSIVE | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | INTERACTIVE EXPERIENCE | | VIDEO STREAMING | INTERACTIVE EXPERIENCE | AUGMENTED REALITY | VIRTUAL REALITY | VIDEO STREAMING | INTERACTIVE EXPERIENCE | AUGMENTED REALITY | VIRTUAL REALITY | |

**ACTIVITY-BASED TIMELINE**

- In-person theatre includes an audience which attend in person, personalised timelines are not usual, though there are several examples … (🚫)
- Video-streaming is usually presented to an audience with a global timeline … (🚫)
- Video-streaming is usually presented to an audience with a global timeline … As Hybrid theatre includes an audience which attend in person, personalised timelines are not usual … (🚫)
- Augmented Reality experiences are premised on the characteristics of the environment that the viewer is in, therefore highly varied environments (i.e. Hybrid theatre) are likely incompatible … As Hybrid theatre includes an audience which attend in person, personalised timelines are not usual … (🚫)
- Video streaming is not an immersive offering, rather it may offer the mechanics of delivering the information to remote audiences participating in immersive theatre (🚫)
- IN SCOPE OF RESEARCH (Immersive Interactive Experience)
- IN SCOPE OF RESEARCH (Immersive Augmented Reality)
- IN SCOPE OF RESEARCH (Immersive Virtual Reality)

**LOCATION BASED / REMOTE**

**GLOBAL TIMELINE LIVE AND ON-DEMAND**

- In-person theatre includes an audience which attend in person. This is one of Stagetexts's core service offerings. (✓)
- Video-streaming is usually presented to an audience with a global timeline. This is one of Stagetexts's core service offerings. (✓)
- IN SCOPE OF RESEARCH (Hybrid Video Streaming)
- IN SCOPE OF RESEARCH (Hybrid Interactive Experience)
- In-person theatre and Augmented Reality experiences are premised on the characteristics of the environment that viewer is in, therefore highly varied environments are likely incompatible … (🚫)
- IN SCOPE OF RESEARCH (Hybrid Virtual Reality)
- Video streaming is not an immersive offering, rather it may offer the mechanics of delivering the information to remote audiences participating in immersive theatre (🚫)
- IN SCOPE OF RESEARCH (Immersive Interactive Experience)
- IN SCOPE OF RESEARCH (Immersive Augmented Reality)
- IN SCOPE OF RESEARCH (Immersive Virtual Reality)

**LOCATION BASED / REMOTE**

Click here for a higher resolution version of this diagram

In our review of information from The ImAc (Immersive Accessibility) project, which focuses on developing and testing solutions for subtitles in 360º videos, we identified the key objectives of the ImAC product include:

- Ensuring that subtitles do not break immersion.
- Assisting users in better comprehending the story.
- Testing rendering modes and guiding methods for effective subtitle placement in immersive environments.

*Key findings from our ImAc project review*

- Floating captions that follow the viewer's gaze improve readability and immersion
- Static captions in virtual environments must be strategically placed to ensure access throughout the experience.
- Dynamic positioning techniques are effective in maintaining both accessibility and immersion.

We found that the BBC R&D whitepaper (WHP 330) provides valuable insights into subtitle placement within virtual environments, including techniques to avoid VR motion sickness related to fixed captions and strategies for optimal subtitle placement to enhance viewer comfort and comprehension.

*Key findings from our BBC R&D Whitepaper review*

- Fixed-position captions can lead to VR motion sickness and discomfort.
- Optimal subtitle placement involves dynamic positioning that adapts to viewer movements.
- Research supports the need for innovative, viewer-centred captioning solutions.

The **gap analysis** identified specific challenges in applying traditional theatre captioning methods to extended reality theatre, such as the need for captions that move dynamically with the audience's perspective. Fixed-position captions are insufficient for maintaining accessibility in 360º settings, as they often fall out of the viewer's field of vision. While elements of the existing business process remain, they need to be augmented to enable captions to adapt

dynamically in real-time to ensure they remain within the viewer's line of sight and do not disrupt the immersive experience and for 'spatial information' captions to co-exist to indicate action out of the user's field of vision or outside of the virtual world.

*Key findings from our Gap Analysis*

- In the current process, there is no way to indicate where a caption should be positioned in relation to the virtual space.
- There is no mechanism to track where the audience member is looking
- There is no approach to ensure captions follow the viewer's gaze and remain within their line of sight
- The current process does not include a method to alert viewers when significant actions occur outside their immediate field of view.
- There is no system to integrate notifications for actions happening outside the virtual world, resulting in potential loss of important context and information.
- Without appropriate notifications, viewers may miss critical elements of the performance, diminishing the overall immersive and narrative experience.

# Concepts in extended reality theatre captioning for deaf audiences

Before delving into our proposal to address these gaps, it's essential to understand the unique challenges presented by extended virtual environments. A thorough grasp of user interaction within these spaces is crucial for designing an effective solution.

Captioning requirements for extended reality performances differ significantly from traditional formats due to the need for captions to dynamically relate to the audience member's field of view and the ongoing action. While many principles from in-person and online captioning apply, this spatial relationship is pivotal and must be fully understood.

The presentation layer for extended reality performances, such as through VR headsets, holds the key information regarding where the audience member is looking. It is responsible for computing and making decisions on how captions are presented.

The key concepts we have identified are:

1. **Optimal Caption Presentation**: Research from BBC Research & Development highlights potential issues with VR sickness caused by static text in a moving virtual environment. This underscores the importance of choosing the most suitable caption presentation methods.
2. **Dynamic Contextual Relevance**: Unlike static captions, those in immersive theatre must dynamically adapt to the user's field of view and the virtual scene's context. Real-time decision-making capabilities within the captioning system are essential for maintaining relevance and accessibility.
3. **World-Referenced Captions**: Captions anchored to specific objects or characters within the virtual world ensure synchronisation with relevant elements, enhancing user comprehension and engagement.
4. **Guidance Mechanisms**: To address instances where users look away from the referenced object, implement guidance mechanisms and additional fixed-position or 'spatial indicator' captions. This ensures critical information remains accessible regardless of the user's focus.
5. **Global Timeline**: In extended reality theatre, a global timeline dictates the timing of all elements—captions, character actions, and narrative progression. This synchronisation provides a cohesive experience for all the audience, who are enjoying a shared experience progressing through the narrative as dictated by the production
6. **Activity-based (Personalised) Timeline**: This concept adapts the performance to the individual audience member's interaction. Whether it's an on-demand virtual experience or one that progresses based on audience interaction, this timeline ensures a personalised and engaging experience.

This graphic illustrates some of these key concepts,
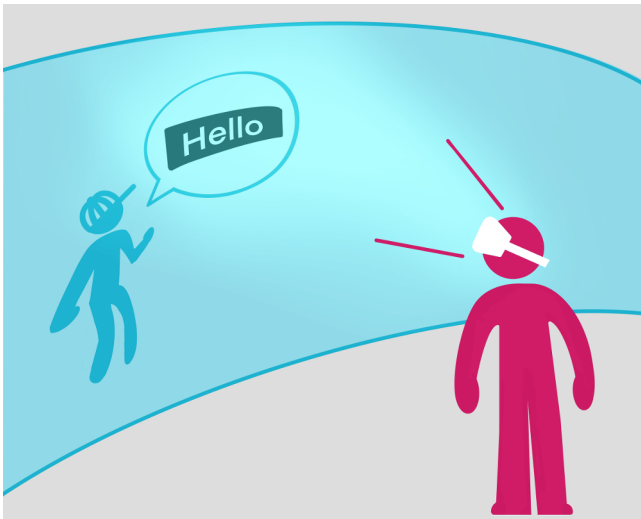


An audience member engrossed, looking at the primary performance activity
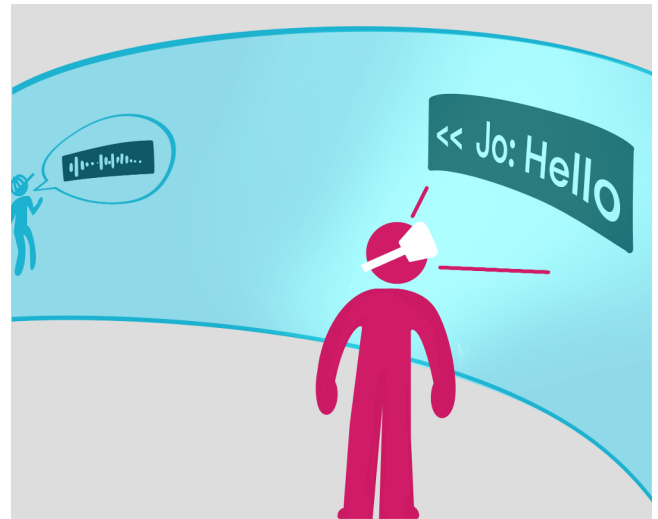


An audience member looking away from the primary performance activity

On the left, you can see an audience member whose immersive field of view is directed towards the stage and the actors on it. The words being spoken by the actor are represented as a caption 'attached' to them - as they speak and move the captions change and move with them. These are 'word referenced' captions, being attached to a person or object within the virtual world.

On the right, you can see an audience member whose immersive field of view is directed away from the stage and the action. While the world-referenced captions are still present, the audience member would not be aware of them - this is where the 'spatial indicator' captions come into play. We can see that the caption is presented in the audience member's field of view with indicators drawing their attention to the caption and the direction the sound is coming from. This way, no information is missed and audience members are prompted that action is happening somewhere other than where they are currently looking.

# Captioning in extended reality performances - critical components

Our research and gap analysis have identified several critical components essential for effective extended reality captioning for deaf audiences. These components form the backbone of a repeatable and scalable captioning process. These are:

- Providing critical information for each line of captions
- Ability to map items in the virtual world to captions

## 1. Critical information for caption lines

We have determined the minimum critical pieces of information to be codified into each caption line in the caption file. These elements ensure that captions are informative, timely, and contextually accurate:

- **WHAT**: The caption text itself. This is the actual dialogue or sound description that needs to be presented to the viewer. Dialogue needs to be verbatim and sound description needs to be factual.
- **WHEN**: The timing for the caption's appearance and disappearance. Accurate timing is crucial for synchronising captions with the global timeline of the performance, ensuring that audience members receive the information at the correct moment. Timing captions perfectly in time with speech supports lipreading.
- **WHO**: The world reference context of the caption. This indicates the object or character associated with the caption, providing essential context for the viewer to understand who or what the caption pertains to. When captions are 'attached' to a person in this way, character names do not need to be included in the caption.

## 2. Object dictionary Integration

An object dictionary serves as a lookup table, mapping voice-span tags in the caption file to objects in the virtual environment. This integration is vital for providing context and ensuring that captions are dynamically positioned within the viewer's field of view.

**Voice span tags**: Voice span tags are metadata embedded within the caption file that identify the source of the dialogue or sound. These tags are used to link specific lines of caption text to corresponding objects, characters, or elements within the virtual environment. For example, a voice span tag might indicate that a particular line of dialogue is spoken by a specific character or originates from a particular location within the scene.

**Mapping voice span tags**: The object dictionary maps voice span tags in the caption file(s) to corresponding objects in the immersive world. This association allows the system to determine the relevant visual element each caption refers to.

**Field of view evaluation**: The application evaluates whether the world reference (the object or character) is within the user's field of view. If the object is outside the field of view, the system can provide localisation (guidance) information to the caption user.

**Spatial captions for guidance**: When an object or character associated with a caption is outside the viewer's immediate field of view, spatial indicator captions are used to provide guidance. These spatial captions indicate the direction or location of the relevant object, helping the viewer to orient themselves and find the source of the captioned information.

# Developing a business process for captioning in extended reality

Using the output of our research, we designed a business process for preparing and presenting captions in extended reality theatre productions, that builds on the existing approach for caption creation but which also includes a framework that combines caption files with object dictionaries. This process supports the creation of a potentially repeatable mechanism for integrating captions into immersive projects.

# Codifying extended reality captions into WebVTT caption files

We have chosen an open standard of caption files as part of our research, when looking at how captions in an immersive environment should be created and displayed. We have determined that one or more caption files are required to enable a captioned, global timeline pre-recorded immersive experience,

- Caption files are created by a trained captioner and are provided to the application developer
- The application developer integrates the caption file into the immersive presentation

Whilst WebVTT supports a definition of HOW each caption could be presented (such as caption styling information, caption placement) and this works well for in person and on-line experiences, these definitions are not likely to  be relevant for an extended reality presentation where,

- In **preparation** the application developer shall define caption display modes
- In **presentation** real time decisions must be made by the application that considers
    - *the user's context including objects within their field of view*
    - *the WHAT, WHEN and WHO (the object) the caption line relates to (if any)*

## Caption lines without world referenced context

```
00:00:03.254 --> 00:00:13.680
MOBILE PHONE TONES

00:00:15.254 --> 00:02:15.254
Would you please take a moment now
to make sure your mobile phone
is switched off.  Thank you.
```

For caption lines that do not reference an object in the immersive space (examples could include narration, background music or emergency notifications) WebVTT caption text body without markup could be defined as per the example on the left

## Caption lines with world referenced context

For caption lines that do reference an object in the immersive space (examples could include lines spoken by an actor, audio emitted from a piece of equipment in the virtual environment) the WebVTT caption text body is to utilise voice span tag elements as per the example on the right to attribute the caption to the object.

```
00:00:03.254 --> 00:00:13.680
<v telephone>RINGING</v>

00:00:13.680 --> 00:00:16.254
<v mark>Why are you not answering that!</v>
```

# Creating a dictionary of objects for global timeline pre-recorded immersive experiences

We believe that one or more object dictionaries are required to enable a captioned, global timeline immersive experience,

- the object dictionary is a lookup table created by the application developer
- this lookup table maps voice span tags in the caption file/s to objects in the immersive world
    - *this enables the application to evaluate whether the world reference is within the user's field of view*
    - *this enables spatial (guidance) information to be provided to the caption user where applicable*

To date, we have not explored potential formats for the Object dictionary - however, it could be,
- A separate configuration item (such as a JSON file) managed as an asset in the immersive project
- Encoded into the caption file, potentially as metadata tracks

# Key Steps in the Business Process

1. **Caption Creation**: A trained captioner develops caption files containing text, timing, and context information.
2. **Object Dictionary Development**: Application developers create a lookup table that links captions to virtual objects.
3. **Integration**: Caption files and object dictionaries are integrated into the immersive application.
4. **Real-Time Decision Making**: A developed application would need to dynamically adjust caption presentation based on user interaction within the virtual environment.

# High level business logic for immersive experience caption presentation

The diagram below describes the real time decision making process we envisage for extended reality presentations,
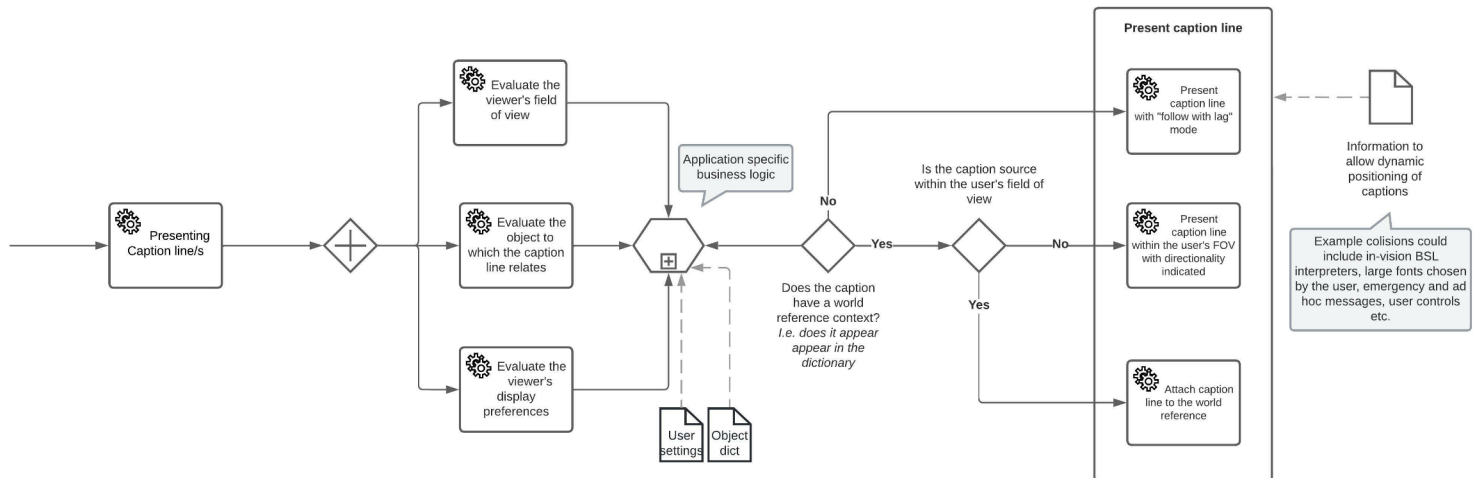


Click here for a higher resolution version of this diagram

## POTENTIAL FOR LOW TECHNOLOGICAL BARRIERS

Our research suggests that the technological barrier to entry for implementing extended reality caption solutions for theatre productions could be low. The integration of caption files with object dictionaries and the adoption of standardised formats such as WebVTT facilitate the development process. This opens the door for wider adoption and innovation in the field of immersive captioning.

# Conclusion

Our research into captioning of extended reality theatre productions for deaf accessibility has led to the development of a robust framework that addresses the unique challenges of virtual environments. By understanding the fundamental concepts and requirements, we have created a scalable and repeatable process that could be developed for integrating captions into extended reality projects. Considering access at the outset of planning an extended reality performance is key.

Contact Stagetext if you would like to discuss how to make your performances accessible for deaf, deafened and hard of hearing audience members, and contact our technology consultants Peak Signal for information on the technical processes proposed.

## About Stagetext

We are Stagetext, a deaf-led charity. We are passionate about making the arts a more welcoming and accessible place.

18 million people in the UK are deaf, deafened or hard of hearing. That means that more than 1 in 3 of us rely on subtitles and captions to experience the joy of live theatre, to sing along with our favourite musicals, to take guided tours around museums, or to see our favourite authors speak at festivals.

We make these experiences possible by providing text-based deaf access to arts and culture through theatre captions, live subtitles and digital subtitles. Using our two decades of experience, we set the highest standards in access, to give everyone a complete and equal experience of the arts.

https://www.stagetext.org/

## About Peak Signal

Peak Signal specialises in providing strategic, delivery, and improvement services to the culture, media and entertainment industry. With a proven track record of successful assignments across strategy, delivery and improvement, the company excels in delivering innovative approaches to provide insightful and practical solutions. Our team draws on broad industry experience and networks to create actionable insights and powerful outcomes for our customers.

https://peaksignal.io/