



ESCUELA DE INGENIERÍA DE FUENLABRADA

GRADO EN INGENIERÍA EN SISTEMAS
AUDIOVISUALES Y MULTIMEDIA

TRABAJO FIN DE GRADO

TÍTULO DEL TRABAJO CON LETRAS MAYÚSCULAS
PARA SUSTANTIVOS Y ADJETIVOS

Autor : Juan José Arias Rojas

Tutor : Dr. Jesús María González Barahona

Curso académico 2024/2025



©2025 Juan José Arias Rojas

Algunos derechos reservados

Este documento se distribuye bajo la licencia

“Atribución-CompartirIgual 4.0 Internacional” de Creative Commons,

disponible en

<https://creativecommons.org/licenses/by-sa/4.0/deed.es>

*Dedicado a
todos aquellos que me animaron y apollaron
incluso en mis momentos de debilidad*

Agradecimientos

Resumen

Summary

Índice general

1. Introducción	1
2. Tecnologías utilizadas	3
2.1. Tecnologías principales	3
2.1.1. A-frame	3
2.1.2. HTML5	6
2.1.3. JavaScript	7
2.1.4. Three.js	8
2.1.5. WebXR	9
2.1.6. WebGL	10
2.2. Tecnologías auxiliares	11
2.2.1. Visual Studio Code	11
2.2.2. Github	11
2.2.3. Meta Quest 3	12
2.2.4. LaTeX	13
3. Desarrollo del proyecto	15
3.1. Sprint 0	15
3.1.1. Objetivo Principal	15
3.1.2. Implementación	16
3.1.3. Resultados	16
3.2. Sprint 1	17
3.2.1. Objetivo Principal	17
3.2.2. Implementación	17

3.2.3. Resultados	21
3.3. Sprint 2	22
3.3.1. Objetivo Principal	22
3.3.2. Implementación	22
3.3.3. Resultados	22
3.4. Sprint 3	22
3.4.1. Objetivo Principal	22
3.4.2. Implementación	22
3.4.3. Resultados	23
3.5. Sprint 4	23
3.5.1. Objetivo Principal	23
3.5.2. Implementación	23
3.5.3. Resultados	23
3.6. Sprint 5	23
3.6.1. Objetivo Principal	23
3.6.2. Implementación	23
3.6.3. Resultados	23
3.7. Sprint 6	24
3.7.1. Objetivo Principal	24
3.7.2. Implementación	24
3.7.3. Resultados	24
4. Resultados	25
5. Pruebas y experimentos	27
6. Conclusiones	29
6.1. Aplicación de lo aprendido	29
6.2. Lecciones aprendidas	29
6.3. Trabajos futuros	29
Bibliografía	31

Índice de figuras

2.1. Escena básica de A-frame	4
2.2. Escena de Three.js	9
2.3. Meta Quest 3	13
3.1. Escena de test de A-Frame	16
3.2. Esquema manos segun WebXR	18
3.3. Primera demo de manos	21

Capítulo 1

Introducción

Capítulo 2

Tecnologías utilizadas

En esta sección se muestran las distintas tecnologías que han sido necesarias para el desarrollo de este proyecto tanto de forma directa como indirecta. Del mismo modo, se explicará su funcionalidad y la manera en la que contribuyeron al proyecto.

2.1. Tecnologías principales

Aquí se nombran y explican aquellas tecnologías que han tenido una implicación directa con el proyecto y que han tenido un papel crucial e imprescindible con las cuales, sin ellas, no se habría podido llegar a los resultados obtenidos. Las principales tecnologías utilizadas han sido HTML5, JavaScript y A-Frame, estas tres tecnologías han sido el núcleo principal del proyecto. El resto de las tecnologías principales son aquellas en las que se apollan esas tres, permitiendo que funcionen como lo hacen.

2.1.1. A-frame

A-Frame [1] es un web framework diseñado para construir experiencias de realidad virtual. A-frame está basado en HTML y en JavaScript, permitiendo realizar cualquier escena que uno pueda imaginar sin necesidad de conocimientos avanzados en gráficos 3D. Al estar basado en HTML es posible realizar escenas simples directamente desde un archivo HTML simplemente importando la librería de A-frame y añadiendo dentro de la etiqueta `<a-scene>` cualquier elemento que desee el usuario.

Uno de los ejemplos más simples que se puede llegar a realizar utilizando A-frame es el siguiente:

```
<html>
  <head>
    <script src="https://aframe.io/releases/1.6.0/aframe.min.js"></script>
  </head>
  <body>
    <a-scene>
      <a-box position="-1 0.5 -3" rotation="0 45 0" color="#4CC3D9"></a-box>
      <a-sphere position="0 1.25 -5" radius="1.25" color="#EF2D5E"></a-sphere>
      <a-cylinder position="1 0.75 -3" radius="0.5" height="1.5" color="#FFC65D"></a-
        cylinder>
      <a-plane position="0 0 -4" rotation="-90 0 0" width="4" height="4" color="#7BC8A4"></a-
        plane>
      <a-sky color="#ECECEC"></a-sky>
    </a-scene>
  </body>
</html>
```

Listing 2.1: Escena A-Frame básica

Este código de HTML genera una de las escenas más básicas que se pueden hacer en A-frame. La cual consiste en un plano y, encima, una serie de figuras geométricas, como se aprecia en la figura 2.1.

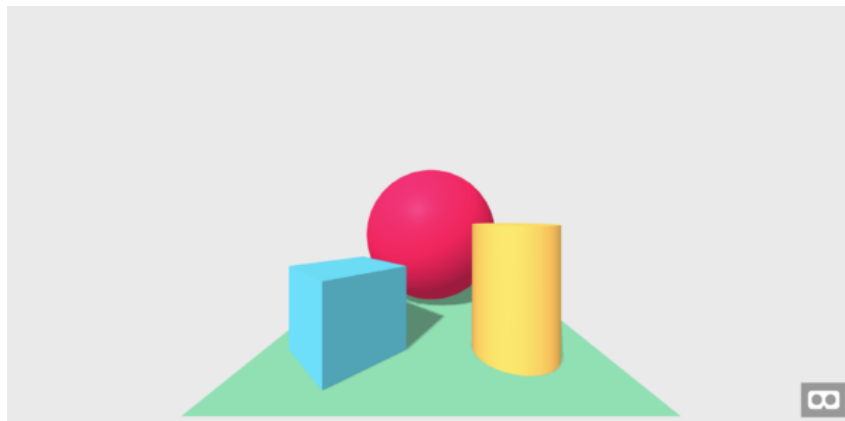


Figura 2.1: Escena básica de A-frame

Una de las principales características de esta tecnología es su arquitectura basada en un modelo de entidad-componente (*Entity Component System*), donde cada objeto dentro de la escena es una entidad que luego es integrada con Three.js para crear la escena. Una entidad en A-frame es un objeto HTML que se crea añadiendo la etiqueta `<a-entity>` y el componente

es la apariencia, comportamiento y funcionalidad que se le asigna mediante JavaScript.

Para hacer uso de un componente se requiere de una serie de pasos previos. Si el componente es externo es necesario importarlo en la cabecera del HTML del mismo modo que se importa A-frame y posteriormente añadirlo a la entidad en la que uno desee usarlo. Por otra parte, si el componente es de nuestra creación, primero hay que registrar y definir el componente. Esto se hace en un archivo externo de JavaScript mediante el método de `A-Frame.registerComponente`. A este método, del mismo modo que una función primero se le asigna un nombre, este será el nombre del componente. Luego, este método consta de varias funciones internas las cuales se encargan de definir la apariencia y lógica de dicho componente. Algunas de estas funciones son la función `schema`, `init` o `tick`. La función `schema`, define las propiedades principales del componente, estas se pueden modificar dependiendo de la lógica del componente desde el HTML a la hora de llamarlo. La función `init` se ejecuta una única vez al inicio cuando se inicializa el componente y la función `tick` se ejecuta a cada frame de la escena. Una vez el componente está registrado y programado importamos el archivo que contiene el componente al HTML y lo insertamos a la escena en la entidad u objeto que deseemos.

Un ejemplo de la llamada de un componente dentro de un archivo de HTML se puede apreciar en 2.2

```
<html>
  <head>
    <script src="https://aframe.io/releases/1.6.0/aframe.min.js"></script>
    <script src = "componente1.js"></script>
    <script src = "componente2.js"></script>
  </head>
  <body>
    <a-scene>
      <a-entity componente1></a-entity>
      <a-box componente2></a-box>
    </a-scene>
  </body>
</html>
```

Listing 2.2: Ejemplo de llamada de entidad

En este ejemplo primero se muestra como a una entidad vacía se le añade un componente llamado `componente1` y también se muestra como a una entidad de cubo se le añade el componente `componente2`. Del mismo modo, para la creación de un componente, en el archivo `.js` que posteriormente importaremos en la escena, únicamente es necesario registrar el

componente y definir sus características como en el siguiente ejemplo:

```
AFRAME.registerComponent('componente1', {  
  init() {  
    const box = document.createElement('a-box');  
    box.setAttribute('position', '-1 0.5 -3');  
    box.setAttribute('rotation', '0 45 0');  
    box.setAttribute('color', '#4CC3D9');  
  
    const sphere = document.createElement('a-sphere');  
    sphere.setAttribute('position', '0 1.25 -5');  
    sphere.setAttribute('radius', '1.25');  
    sphere.setAttribute('color', '#EF2D5E');  
  
    this.el.appendChild(box);  
    this.el.appendChild(sphere);  
  }  
});
```

Listing 2.3: Ejemplo de creacion de entidad

Este fragmento de código registra el componente `componente1` el cual crea un cubo y un cilindro con unos colores, posición y rotación específicos y los añade a la escena.

Para este proyecto, A-Frame ha sido la tecnología principal, permitiendo elementos fundamentales como la creación y gestión de entornos de realidad virtual. También, permitio la integración de componentes diseñados mediante JavaScript y permitio el manejo de interacciones mas complejas a la vez que optimizaba el rendimiento de la escena 3D.

2.1.2. HTML5

HTML5 [3] es la quinta iteración del lenguaje de HTML (*marcado de hipertexto*), el cual es la estructura básica y principal de toda página web. Desarrollado conjuntamente por W3C (*World Wide Web Consortium*) [8] y WHATWG (*Web Hypertext Application Technology Working Group*) [9]. HTML5 introduce mejoras sustanciales respecto a sus anteriores versiones, incluyendo nuevas etiquetas semánticas, soporte multimedia mejorado y compatibilidad ampliada con diversos navegadores y dispositivos, dándole mayor flexibilidad y dinamismo a la creación de páginas web.

Una de sus nuevas introducciones clave es la introducción de etiquetas semánticas como: `<header>`, `<article>`, `<section>` y `<footer>`, que facilitan una estructuración clara y accesible del contenido web. Estas etiquetas no solo ayudan a mejorar la organización de la

información, si no que también facilitan la búsqueda de información por parte de los motores de búsqueda y facilitan la interpretación por parte de los desarrolladores. Además, HTML5 incorpora nuevas interfaces de programación de aplicaciones (*API*), destacándose las funcionalidades de almacenamiento local mediante `localStorage` y `sessionStorage`, que permiten la gestión eficiente de datos directamente en el navegador, eliminando la necesidad de bases de datos externas.

HTML5 también introduce mejoras significativas en la gestión de contenido multimedia, eliminando la necesidad de complementos externos. Para esto se implementaron las etiquetas `<audio>` y `<video>` las cuales permiten la incorporación directa de archivos de sonido y video en las páginas web, permitiendo mayor dinamismo en el desarrollo. Otro elemento que se añadió fue el elemento `<canvas>` el cual habilita la generación de gráficos y animaciones en tiempo real a través de JavaScript, permitiendo el desarrollo de aplicaciones interactivas y videojuegos en línea.

HTML5 se ha establecido firmemente como un estándar en el desarrollo de aplicaciones web progresivas (*PWA*) y en entornos multiplataforma. Su integración con tecnologías como CSS3 y JavaScript permite la creación de interfaces dinámicas y adaptativas, compatibles con una amplia gama de dispositivos, desde ordenadores hasta tabletas y teléfonos.

Gracias a que HTML es una tecnología extensible, tecnologías como A-Frame o WebXR han sido posibles de utilizar ya que son extensiones de HTML. WebXR permitió poder trabajar con entornos VR desde el navegador mientras que A-Frame fue el eje central del proyecto, donde se estructuró la escena. HTML fue una de las tecnologías más importantes ya que ofrece una estructura esencial para el desarrollo del proyecto, haciendo uso de sus extensiones de A-Frame y WebXR para poder mostrar los elementos de la escena en la realidad virtual.

2.1.3. JavaScript

JavaScript es un lenguaje de programación ligero y multiplataforma utilizado principalmente para la creación de contenido dinámico e interactivo para páginas web. Su flexibilidad permite desarrollar elementos para mejorar la interacción del usuario en páginas web tanto del lado del servidor como del lado del cliente. En 1997, JavaScript fue estandarizado por *ECMA* como ECMAScript y poco después como un estándar *ISO*.

JavaScript, creado por Brendan Eich de Netscape en 1995 bajo el nombre de *Mocha*, pos-

teriormente se renombró a LiveScript y finalmente a JavaScript. En el año 2000, JavaScript se extendió al lado de los servidores con la introducción de tecnologías como *Node.js*, y posteriormente su popularidad creció con la llegada de *AJAX*. Y Con la llegada de ECMAScript 6 en 2015, se introdujeron características mas avanzadas y actualizaciones anuales, haciendo que hoy en día sea uno de los lenguajes mas importantes en el desarrollo web.

JavaScript es un lenguaje de programación de alto nivel, lo que significa que su lenguaje esta diseñado para ser lo mas 'humano' posible, permitiendo una rapida comprensión del codigo sin necesidad de un nivel alto de conocimientos. Es un lenguaje basado en eventos ya sean entradas de ratón o entradas de teclado, permitiendo la creación de interfaces de usuario interactivas. JavaScript puede integrarse en un HTML dentro de la etiqueta `<script>` de forma directa, escribiendo el codigo. También, se puede introducir codigos de JavaScript, los cuales estén en archivos distintos con la extensión correspondiente al lenguaje (.js), esto es posible vinculando dicho archivo en la cabecera del HTML.

Para este proyecto formo un papel crucial, ya que fue con JavaScript que se desarrollaron los distintos componentes de A-Frame que se usaron para el funcionamiento del proyecto. Permitio diseñar e implementar la lógica que hay tras cada componente permitiendo alcanzar los resultados obtenidos.

2.1.4. Three.js

Three.js [10] es una biblioteca de codigo abierto de JavaScript utilizada para la creación de gráficos 3D en los navegadores web. Three.js funciona sobre WebGL [2], la cual permite generar y visualizar animaciones y escenas 3D en el navegador sin necesidad de complementos. La tecnologia de WebGL tambiénpuede usarse junto con el elemento `<canva>` de HTML.

Esta biblioteca proporciona una API de alto nivel la cual permite al usuario la creación y manipulación de geometrías 3D como cubos, esfera o planos, así como la aplicación de texturas o la aplicación de tanto camaras como de efectos de iluminación en las escena, permitiendo que el desarrollo de dichas escenas sea mucho mas simplificado. También, Three.js permite la posibilidad de importar modelos 3D desde archivos distintos creados desde aplicaciones distintas.

La biblioteca de Three.js ofrece una alta variedad de herramientas que permiten la gestión y control de usuario, facilitando asi tanto la navegación como la interacción dentro de las escenas 3D. También, Three.js soporta animaciones suaves y dinámicas tanto para objetos en la

escena como para camaras, lo cual permite la creación de efectos visuales mas dinámicos e impresionantes al igual que juegos interactivos.

Gracias a la amplia variedad de posibilidades que permite la biblioteca de Three.js las opciones de los elementos o escenas que se pueden llegar a crear son basicamente ilimitadas. Siendo capaz de crear escenas complejas como la de la figura 2.2



Figura 2.2: Escena de Three.js

Three.js permitio añadir una capa mas simplificada a WebGL, simplificando asi el proceso de creación y manipulación de los elementos 3D. Permitió crear un entorno 3D con mayor dinamismo y control para el proyecto.

2.1.5. WebXR

WebXR [7] es una tecnología desarrollada por W3C (*World Wide Web Consortium*) que permite crear experiencias inmersivas desde el navegador. Esta tecnología combina la Realidad Aumentada (AR) y la Realidad Virtual (VR) con la accesibilidad de un navegador web. Lo cual permite a los usuarios experimentar e interactuar con entornos tridimensionales a través de cualquier dispositivo con acceso a un navegador compatible.

A medida que la tecnología de WebXR ha ido evolucionando, se han desarrollado distintos tipos de experiencias para poder adaptarse a distintos contextos y necesidades. Hoy en día, se podria clasificar los distintos tipos en 3 categorias:

- **WebXR AR:** Este tipo de WebXR combina el mundo virtual y el mundo real, permitiendo que distintos elementos del mundo virtual puedan superponerse en el entorno físico a

través de la cámara de un dispositivo compatible, pero sin que llegue a influir el mundo real en los elementos virtuales.

- **WebXR VR:** Este modo permite a los usuarios sumergirse en el entorno virtual creado, y con la ayuda de dispositivos suplementarios como auriculares, la experiencia es aún mayor. Esta tecnología permite a los usuarios interactuar y experimentar en primera persona distintos entornos virtuales, desde juegos y entretenimiento, hasta simulaciones virtuales y visualizaciones en 3D.
- **WebXR MR(Mixed Reality):** Esta tecnología combina el mundo real y el virtual de forma mas profunda que la tecnología AR. Permite a los usuarios poder interactuar con elementos tanto virtuales como físicos y que estos puedan interactuar entre si. Este tipo de combinación proporciona un nivel mayor de interacción entre el usuario y su entorno, dando mayor numero de posibilidades para la creación de contenido.

Estos distintos tipos de WebXR ofrecen distintos tipos de experiencias dependiendo del entorno virtual que se quiera diseñar.

2.1.6. WebGL

WebGL (*Web graphics Library*), se trata de una tecnología de bajo nivel multiplataforma usada para la renderización de gráficos tanto tridimensionales como bidimensionales dentro de cualquier navegador que sea compatible.

WebGL fue lanzado en 2011 por el grupo Khronos. Esta tecnología se fundamenta en OpenGL ES, la cual es una variante simplificada de OpenGL, diseñada para dispositivos móviles. WebGL ha sufrido numerables actualizaciones, lo cual ha permitido una evolución continua que ha ido mejorando tanto su funcionalidad como compatibilidad tanto en navegadores de escritorio como en navegadores en dispositivos móviles.

WebGL esta diseñado para trabajar directamente con la GPU (*Graphic Processing Unit*) del dispositivo, lo cual permite un mayor aprovechamiento de la computación para poder generar gráficos detallados y de alta calidad. Además, la tecnología de WebGL esta diseñada para poder integrarse de manera fluida con otros estandares de desarrollo web como HTML, CSS o DOM.

2.2. Tecnologías auxiliares

En esta sección se detallan y describen las distintas tecnologías que aunque no hayan influido de manera directa en los resultados del proyecto, han tenido un papel crucial para el desarrollo de este.

2.2.1. Visual Studio Code

Visual Studio Code (*VS Code*) es un potente editor de código abierto el cual esta disponible para Windows, Linux, MacOS y versión web. VS Code es una de las plataformas de edición de código más utilizadas a nivel mundial por toda clase de desarrolladores de software debido a su versatilidad, flexibilidad y amplia gama de extensiones que facilitan la edición y depuración de código.

También, gracias a su pestaña de **Source Control**, permite ir guardando el código en GitHub para tener un control de las distintas versiones del código.

Gracias a las numerosas extensiones que hay disponibles facilitan en gran medida la creación y depuración de código, permitiendo cosas como autocompletar palabras o expresiones o permitir compilar o depurar con algun lenguaje que VS Code no tenga por defecto. Algunas de las extensiones utilizadas en este proyecto han sido:

- **A-frame completion:** Permite autocompletar rapidamente a la hora de escribir los distintos elementos o Snippets que posee A-frame
- **Error lens:** A la hora de depuración, permite visualizar dentro del código si hay algun error de sintaxis o lógica
- **LaTeX Workshop:** Permite la escritura, compilación y previsualización de la memoria de este proyecto.

2.2.2. Github

GitHub es una plataforma basada en la nube que permite almacenar distintos repositorios y en cada repositorio código. Github esta basado en Git [4], el cual fue creado en 2005 por Linus Torvalds como un sistema de control de versiones de código abierto. Este sistema fue

desarrollado con la intención de ayudar a los desarrolladores a tener en cualquier dispositivo con acceso a internet todo el historial de código que están desarrollando.

Esta tecnología es ampliamente utilizada a nivel mundial por desarrolladores de todo tipo debido a su capacidad de almacenar, soportar y gestionar proyectos de toda clase. Una de las funciones clave que presenta la plataforma, es la capacidad de crear ramificaciones (branches). Esta opción permite a los desarrolladores poder crear copias del código del repositorio en el que están trabajando para poder trabajar en paralelo y posteriormente poder integrar los cambios realizados en paralelo al código principal. Esta opción también permite que más de un desarrollador pueda trabajar en el mismo código, haciendo que cada uno trabaje en paralelo en ramas distintas.

GitHub aprovecha todas las ventajas que permite Git, permitiendo crear repositorios en la nube para sus proyectos Git, que los desarrolladores pueden llegar a compartir con otras personas. Esta plataforma también facilita la colaboración entre desarrolladores y aparte de las ventajas ya mencionadas de Git permite también el uso de herramientas de gestión de proyectos al igual que acciones automáticas.

2.2.3. Meta Quest 3

Meta Quest se trata de un dispositivo inalámbrico diseñado para disfrutar de contenido de realidad virtual (VR) por Meta [6]. Dicho dispositivo está compuesto por unas gafas, un micrófono y auriculares integrados en un único dispositivo que el usuario pone en su cabeza y un par de mandos inalámbricos. La primera versión fue lanzada en 2020 con las Meta Quest 2 y 3 años más tarde, en 2023 se lanzaron las Meta Quest 3.

Para este proyecto se han utilizado unas Meta Quest 3, las cuales cuentan con una memoria de 8 GB de memoria RAM, una resolución de 2064x2208 píxeles por cada uno de los ojos. También, cuenta con mejoras respecto al campo de visión (FOV) respecto a su versión anterior. Para este proyecto jugaron un papel crucial ya que al enfocarse en la detección de mandos el proyecto, sin las gafas no se habría podido probar los resultados que se iban obteniendo.



Figura 2.3: Meta Quest 3

2.2.4. LaTeX

LaTeX [5] es un software de uso libre de creación de documentación de alta calidad. Este sistema es altamente utilizado para la creación de documentación técnica o científica de media o larga extensión, aunque gracias a su versatilidad se puede utilizar para cualquier tipo de documentación.

Entre las distintas características de LaTeX cabe destacar su capacidad de manejar expresiones matemáticas complejas, lo cual es una herramienta indispensable para cualquier documento científico. También gracias a su software motor TeX LaTeX es capaz de convertir los comandos de texto, utilizados para expresar los resultados tipográficos, en un archivo PDF profesional.

Además de facilitar la creación de fórmulas matemáticas complejas, LaTeX también ofrece otras facilidades avanzadas, como puede ser la creación de un índice del contenido, un índice de las figuras utilizadas, gestión de bibliografías o referencias simplificando la organización del documento y la citación de figuras o elementos externos en la bibliografía.

Para este proyecto, LaTeX facilitó considerablemente el trabajo a la hora de crear este documento, ayudando con la creación y gestión de índices, imágenes y bibliografía al igual que a estructurar el texto del documento de forma correcta.

Capítulo 3

Desarrollo del proyecto

En el siguiente capítulo se describe de forma detallada como fue el desarrollo del proyecto. Dicho desarrollo se describe en forma de sprints, desde los inicios y planteamiento del proyecto hasta obtener los resultados finales.

3.1. Sprint 0

Puesta en marcha del proyecto, planteamiento con el tutor y dominio del uso basico de A-Frame.

3.1.1. Objetivo Principal

El objetivo principal de este sprint se puede dividir en 2 partes principales.

- **El planteamiento con el tutor**, donde se identificaron las necesidades para el proyecto del mismo modo que planificar las distintas etapas y duración de cada una y los objetivos que se deseaban obtener.
- **Dominio del uso basico de A-Frame**. Aprender a usar de forma correcta como funcionan las escenas al igual que la creación e implementacion de componentes o el uso de librerias de A-Frame.

3.1.2. Implementación

La identificación de las necesidades y requisitos de un proyecto es una parte fundamental. Durante las primeras charlas con el tutor se concretaron los distintos objetivos y los posibles caminos que puede tomar este proyecto. Al principio se plantearon varios caminos, pero la idea general de todos ellos era la creación de un sistema de handtracking que sea capaz de interactuar con distintos elementos dentro de la escena.

Para lograr llegar a ese objetivo primero fue necesario dominar los elementos básicos de A-Frame como bien a ser la organización de la escena, el uso y creación de componentes y el uso de eventos. Para ello, primero se creó primero el repositorio de *Git*¹ con el que trabajaremos durante todo el proyecto. Se crearon una serie de escenas sencillas con el objetivo de dominar el uso de componentes y eventos simples.

3.1.3. Resultados

Para familiarizarme aun mas con el entorno de A-FRAME se crearon una serie de escenas muy sencillas donde se creaba y utilizaba un componente customizado y se hacía uso del evento nativo de A-Frame `click`. Estas escenas permitieron un mayor entendimiento del funcionamiento tanto de componentes como de eventos, que fueron fundamentales para el resto del proyecto. El resultado de dichas pruebas para mejorar la comprensión de A-Frame se pueden apreciar en las figura 3.1.

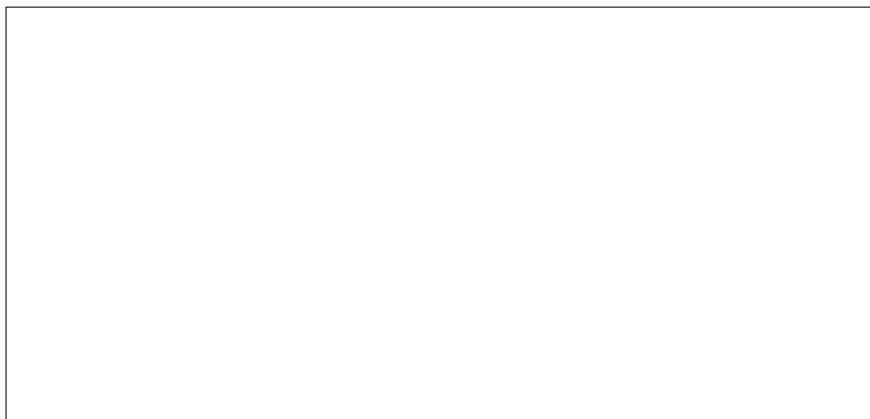


Figura 3.1: Escena de test de A-Frame

¹<https://github.com/JuJoarias/TFG>

Tras la familiarización con el entorno de A-Frame y sus características el proyecto fue formalizado formalmente. Tras ellos se planteo el siguiente sprint del proyecto donde ya se empezarian los primeros pasos formales de este.

3.2. Sprint 1

Investigando el handtraking

3.2.1. Objetivo Principal

En este primer sprint oficial del proyecto, el objetivo princpal era investigar y comprender el uso del handtraking en A-Frame. Este era el paso principal en el que se basaria todo el proyecto, ya que sin una comprensión del funcionamiento del handtracking en A-Frame este proyecto habria sido imposible de realizar.

3.2.2. Implementación

Los primeros pasos fue investigar el handtracking en A-Frame con los elementos ya existentes como *super hands*², o los componentes nativos de A-Frame `hand-controls` o `hand-tracking-controls`. Para ellos se usaron pequeñas escenas de prueba para comprender mejor el comportamiento de las manos en la realidad virtual, y del mismo modo, tener una idea mas clara del objetivo final del aspecto de handtracking.

Tras la investigación de las manos y la sugerencia del tutor, se decidió utilizar la tecnología de WebXR para el handtracking puesto que esta tecnología ya posee un sistema preparado para ello. Además, usando la tecnología de WebXR nos permitia trabajar con cada una de las articulaciones de la mano de forma independiente.

Siguiendo la documentación de *WebXR*³, se empezo el desarrollo de las manos. para ello en el codigo fue necesario tratar cada una de las manos de forma independiente, del mismo modo que cada una de las articulaciones como se muestra en el esquema de las manos de WebXR en la figura 3.2.

²<https://github.com/c-frame/aframe-super-hands-component>

³<https://www.w3.org/TR/webxr-hand-input-1/>

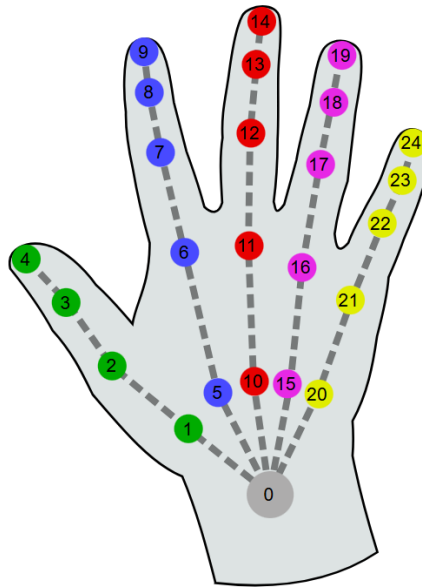


Figura 3.2: Esquema manos segun WebXR

Para poder crear las manos de forma exitosa dentro del entorno de realidad virtual siguiendo WebXR, primero se crea el componente que sera el encargado de detectar y dibujar las manos en este proyecto, dicho componente se llama `hand-skeleton`. Dicho componente no tenia un schema, en el init se definian las distintas variables que eran necesarias para el funcionamiento del código. En el HTML que define la escena unicamente era necesario añadir una vez el componente como una entidad. Pero era imprescindible definir todas y cada una de las articulaciones, como se muestra en el fragmento de código 3.1, pero esa definición unicamente se realizaba en el código de JavaScript.

```
const orderedJoints = [
  ["wrist"],
  ["thumb-metacarpal", "thumb-phalanx-proximal", "thumb-phalanx-distal", "thumb-tip"],
  ["index-finger-metacarpal", "index-finger-phalanx-proximal", "index-finger-phalanx-intermediate", "index-finger-phalanx-distal", "index-finger-tip"],
  ["middle-finger-metacarpal", "middle-finger-phalanx-proximal", "middle-finger-phalanx-intermediate", "middle-finger-phalanx-distal", "middle-finger-tip"],
  ["ring-finger-metacarpal", "ring-finger-phalanx-proximal", "ring-finger-phalanx-intermediate", "ring-finger-phalanx-distal", "ring-finger-tip"],
  ["pinky-finger-metacarpal", "pinky-finger-phalanx-proximal", "pinky-finger-phalanx-intermediate", "pinky-finger-phalanx-distal", "pinky-finger-tip"]
];
```

Listing 3.1: Definición de articulaciones

Después de definir todas las articulaciones era necesario procesarlas. Para ello, se creo la función que se muestra en el fragmento de código 3.2. En la función `renderHandSkeleton` se crea un bucle de código el cual, toma la sesión XR de la escena y por cada input de la sesión toma cada una de las articulaciones, las detecta en la mano y crea una esfera a traves de una segunda función unicamente encargada de crear esferas. Dichas esferas se crean en la posición correspondiente a la mano real y tambien se ajusta el radio de la esfera para que coincida con la mano.

```
renderHandSkeleton: function () {
  const session = this.el.sceneEl.renderer.xr.getSession();
  if (!session || !this.frame || !this.referenceSpace) {
    return;
  }
  const inputSources = session.inputSources;
  for (const inputSource of inputSources) {
    if (inputSource.hand) {
      const hand = inputSource.hand;
      const handedness = inputSource.handedness; // Determina si es la mano derecha o
      izquierda
      for (const finger of orderedJoints) {
        for (const jointName of finger) {
          const joint = hand.get(jointName);
          if (joint) {
            const jointPose = this.frame.getJointPose(joint, this.referenceSpace);
            if (jointPose) {
              const position = jointPose.transform.position;
              if (!this.spheres[handedness + '_' + jointName]) {
                this.spheres[handedness + '_' + jointName] = this.drawSphere(
                  jointPose.radius, position);
              } else {
                this.spheres[handedness + '_' + jointName].object3D.position.
                  set(position.x, position.y, position.z);
              }
            }
          }
        }
      }
    }
  }
},
```

Listing 3.2: Dibujo de las manos en la escena

Esta función se encuentra dentro de la función Tick del componente manos, así que a cada

frame el propio código revisa la posición de todas y cada una de las articulaciones para cada una de las dos manos de entrada. Además de detectar y dibujar las manos en la escena, en esta primera demo se implementó el gesto de Pinch, distinguiendo cada mano y realizando una acción distinta dependiendo de la mano que lo realizase.

Primero, para la detección del gesto se creo la función de `detectGestures`, como se muestra en el fragmento de código 3.3. En dicha función, se tomaba la posición del la punta del dedo índice y del pulgar y se calculaba la distancia entre ambos puntos.

```
detectGestures: function () {
  const session = this.el.sceneEl.renderer.xr.getSession();
  const inputSources = session.inputSources;
  let rightPinching = false;
  let leftPinching = false;
  for (const inputSource of inputSources) {
    if (inputSource.hand) {
      const thumbTip = this.frame.getJointPose(inputSource.hand.get("thumb-tip"), this.referenceSpace);
      const indexTip = this.frame.getJointPose(inputSource.hand.get("index-finger-tip"), this.referenceSpace);
      if (thumbTip && indexTip) {
        const distance = calculateDistance(thumbTip.transform.position, indexTip.transform.position);
        if (distance < pinchDistance) {
          if (inputSource.handedness === 'right') {
            rightPinching = true;
            if (!this.rightPinching) {
              this.createBoxAtHand(inputSource.hand);
            }
          } else if (inputSource.handedness === 'left') {
            leftPinching = true;
            if (!this.leftPinching) {
              this.selectObjectAtHand(inputSource.hand); // Selecciona un objeto cercano para mover
            }
          }
        }
      }
    }
  }
  this.rightPinching = rightPinching;
  this.leftPinching = leftPinching;
},
```

Listing 3.3: Función `detectGestures`

Si la distancia era inferior a 0.02 (valor obtenido tras realizar varias pruebas), se comprobaba que mano era la que lo realizaba y luego se llamaba a la función correspondiente. Si se trataba de la mano derecha, se creaba un cubo verde en la posición de la punta del dedo índice. Y si se trataba de la mano izquierda, si esta se encontraba lo bastante cerca de uno de los cubos creados por la mano derecha (a una distancia inferior a 0.15) la posición del cubo copiaría la de la punta del dedo índice izquierdo hasta que se soltase el gesto.

3.2.3. Resultados

Como resultado de este sprint se creo la prueba de *pinch test*⁴ al igual que en la figura 3.3.

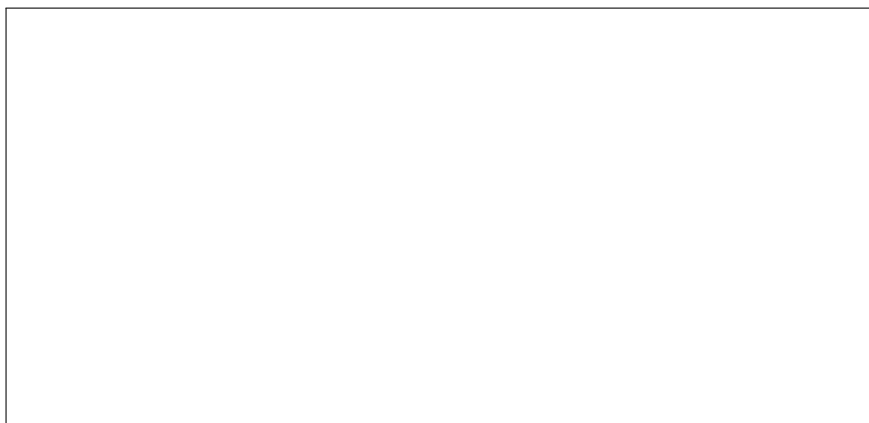


Figura 3.3: Primera demo de manos

Las esferas se dibujaban correctamente y los gestos funcionaban, pero había bastantes aspectos a mejorar. Primero, por sugerencia del tutor, se debía modificar el componente para que fuese necesario añadirlo 2 veces a la escena, 1 por cada mano. También, respecto al gesto, con la mano derecha había que cambiar el funcionamiento ya que el que creara un cubo no sería necesario más adelante. Y respecto a la mano izquierda, se identificó que a la hora de tomar el cubo, este se transportaba hacia la posición del dedo índice. Y aunque esto era lo deseado de primeras, visualmente no quedaba bien ya que si se agarraba una esquina el cubo se transportaba, tragándose la mano en vez de agarrar el cubo mientras mantenía la distancia entre el dedo y el centro del cubo. Aparte de eso, en el caso de que hubiese más de un cubo muy pegado,

⁴https://github.com/JuJoarias/TFG/blob/main/first_steps/first_hands/pinch_test.html

el código no identificaba siempre de forma correcta el mas cercano, haciendo que a veces se tomara el cubo no deseado.

Todos estos factores dieron paso al siguiente sprint, donde se corregirian los errores resalta-
dos. Aparte se intentaria optimmizar la creación de las manos y se añadiría la emisión y gestión
de eventos, ya que en esta primera demo, al realizar el gesto, se llama a la función correspon-
diente si utilizar eventos.

3.3. Sprint 2

Primeros gestos

3.3.1. Objetivo Principal

texto pendiente

3.3.2. Implementación

texto pendiente

3.3.3. Resultados

texto pendiente

3.4. Sprint 3

Nuevos gestos de las manos

3.4.1. Objetivo Principal

texto pendiente

3.4.2. Implementación

texto pendiente

3.4.3. Resultados

texto pendiente

3.5. Sprint 4

Coliders para interacciones

3.5.1. Objetivo Principal

texto pendiente

3.5.2. Implementación

texto pendiente

3.5.3. Resultados

texto pendiente

3.6. Sprint 5

Componentes de cada gesto

3.6.1. Objetivo Principal

texto pendiente

3.6.2. Implementación

texto pendiente

3.6.3. Resultados

texto pendiente

3.7. Sprint 6

Manos finales

3.7.1. Objetivo Principal

texto pendiente

3.7.2. Implementación

texto pendiente

3.7.3. Resultados

texto pendiente

Capítulo 4

Resultados

Capítulo 5

Pruebas y experimentos

Capítulo 6

Conclusiones

6.1. Aplicación de lo aprendido

6.2. Lecciones aprendidas

6.3. Trabajos futuros

Bibliografía

[1] A-Frame. A-frame - introduction, 2024.

<https://aframe.io/docs/1.6.0/introduction/>.

[2] EncodeBiz. Parte 1: WebGL y su impacto en el desarrollo web moderno, 2024.

<https://www.encodebiz.com/blog/parte-1-webgl-y-su-impacto-en-el-desarrollo-web-moderno/cG9zdDozMjc=>.

[3] E. Freeman and E. Robson. *Head First HTML and CSS*. O'Reilly Media, 2nd edition, 2018.

[4] Git SCM. Branching and merging, 2024.

<https://git-scm.com/about/branching-and-merging>.

[5] LaTeX Project. About the latex project, 2025.

<https://www.latex-project.org>.

[6] Meta. Company information, 2024.

<https://www.meta.com/es-es/about/company-info/>.

[7] Onirix. Webxr: Ejemplos y desarrollo en realidad extendida, 2024.

<https://www.onirix.com/es/webxr-ejemplos-desarrollo-realidad-extendida>.

[8] W3C. The w3c mission, 2024.

<https://www.w3.org/mission/>.

[9] WHATWG. Whatwg - web hypertext application technology working group, 2024.

<https://whatwg.org/>.

[10] Wikipedia contributors. Three.js, 2024.

<https://es.wikipedia.org/wiki/Three.js>.