



UNIVERSIDAD PEDAGÓGICA  
NACIONAL

*Educadora de educadores*

# DESARROLLO DE UNA APLICACIÓN EN REALIDAD VIRTUAL PARA MEZCLA DE SONIDO ESPACIAL.

Michael Leonardo Franco Rojas

Universidad Pedagógica Nacional de Colombia.

Facultad de ciencia y tecnología.

Departamento de tecnología.

Licenciatura en electrónica.

Bogotá D.C.

2020

DESARROLLO DE UNA APLICACIÓN EN REALIDAD  
VIRTUAL PARA MEZCLA DE SONIDO ESPACIAL.

Trabajo de grado para optar al título de  
Licenciado en Electrónica.

Autor:

Michael Leonardo Franco Rojas.

Director:

Mg. Diego Mauricio Rivera Pinzón.

Universidad Pedagógica Nacional de Colombia.

Facultad de ciencia y tecnología.

Departamento de tecnología.

Licenciatura en electrónica.

Bogotá D.C.

2020

Nota de Aceptación.

---

---

---

---

Firma del director.

---

Firma del primer jurado.

---

Firma del segundo jurado.

Bogotá, febrero de 2021.

**“Travel to exotic places, avoid the thought”**

Jaron Lanier

*Dedicado a mi madre Ana Franco y a la memoria de mi abuelita Ana Rojas, por su eterno  
amor y compañía.*

*A ellas todos mis logros.*

*A la vida, por hacer que la visión y la audición,  
sean un fenómeno de admirar.*

*Como Educadores, nuestra responsabilidad con las personas,  
es hacer que la realidad sea la mejor aula para aprender,  
y que la virtualidad nunca, por ningún motivo, sea una jaula para enseñar.*

**“Decidirse a no aceptar como verdadero más que aquello que, por su claridad y  
distinción, aparece a la razón como evidente, da al hombre razonable la garantía de  
no errar”**

Rene Descartes

## Agradecimientos

*Nada de lo aquí plasmado se hubiera podido realizar sin el apoyo incondicional de mi madre y mi compañera. A ellas les debo lo que soy y lo que seré, por siempre gracias.*

*Agradecimiento especial a mi madrina y maestra Dora Gama, su sabiduría y experiencia han ayudado para encontrar las respuestas correctas.*

*Agradecimiento especial al profesor Diego Mauricio Rivera Pinzón. Sus conocimientos y enseñanzas llevaron a buen término este proyecto.*

*A la Universidad Pedagógica Nacional y todo el cuerpo docente, gracias por haber transformado mi vida.*

*A mis amigos y colaboradores, por su paciencia y comprensión.*

## Índice de contenido

1	Introducción.....	5
	1.1 Problema de investigación.....	5
	1.2 Objetivos de la investigación .....	6
	1.2.1 Objetivo general.....	6
	1.2.2 Objetivos específicos .....	7
	1.3 Justificación de la investigación.....	7
	1.4 Producto final entregable.....	8
2	Antecedentes y Marco Teórico.....	9
	2.1 Antecedentes .....	9
	2.2 Marco Teórico - Conceptual.....	12
	2.2.1 Sonido y Señales.....	13
	2.2.1.1 Naturaleza del sonido.....	13
	2.2.1.2 El sistema auditivo humano .....	18
	2.2.1.3 Percepción espacial del sonido y audio binaural.....	25
	2.2.1.4 Análisis de sistemas .....	30
	2.2.1.5 Funciones de Transferencia Relacionadas con la Cabeza (HRTF).....	32
	2.2.2 Realidad Extendida.....	36
	2.2.2.1 Realidad Virtual .....	39
	2.2.2.1.1 Oculus Quest .....	41
	2.2.3 Desarrollo de software .....	42
	2.2.3.1 Motor de Videojuego .....	42
	2.2.3.1.1 Unity 3D.....	44
	2.2.3.2 Visual Studio y C# .....	46
	2.2.3.3 SDK de Audio 3D y Ambisonics .....	47

2.2.4	Mezclador Visual .....	49
2.2.4.1	Imaginación.....	49
2.2.4.2	Estéreo y mono.....	50
2.2.4.3	La idea tradicional de los controles.....	52
2.2.4.4	La consola de mezcla .....	52
2.2.4.5	Estación de trabajo de audio digital (DAW) .....	54
2.2.4.6	Conceptos básicos de un mezclador visual .....	56
2.2.4.7	El arte de mezclar.....	57
3	Metodología.....	59
3.1	Modelo de prototipo .....	60
3.2	Etapas metodológicas .....	61
4	Desarrollo .....	64
4.1	Herramientas utilizadas .....	64
4.2	Características de los Recursos .....	64
4.3	Assets y scripts de terceros.....	64
4.4	Assets y scripts propios .....	69
4.4.1	Audio Manager .....	69
4.4.2	Audio Player .....	70
4.4.3	Script reproductor de video.....	71
4.4.4	Script para agarrar objetos .....	72
4.4.5	Otros scripts .....	72
5	Resultados.....	73
5.1	Descripción general.....	73
5.2	Escenarios e interfaz de usuario .....	74
5.2.1	Cuarto de control .....	74

5.2.2	Cuarto de mezcla .....	76
5.2.3	Representación de los sonidos .....	77
5.2.4	Aula.....	78
5.2.5	Menús de mano y menús móviles.....	79
5.3	Mecánica de la aplicación .....	81
5.3.1	Sistema de navegación.....	81
5.3.2	Sistema de mezcla.....	82
5.3.2.1	Mezclador.....	82
5.3.3	Grabación y reproducción.....	84
5.3.4	Reproducción de video clips.....	85
5.3.5	Simulación del concepto de audio 3D .....	85
5.3.6	Traslación y Rotación de los Objetos .....	86
6	Conclusiones.....	87
7	Bibliografía.....	90
8	Anexos .....	94
8.1	Anexo 1 Logotipo.....	94
8.2	Anexo 2: guías de usuario parte 1. ....	95
8.3	Anexo 3: guías de usuario parte 2. ....	96
8.4	Anexo 4: listado de Assets y paquetes .....	97

## Lista de Figuras

Figura 1 Forma sinusoidal de las zonas de rarefacción y compresión de las partículas del aire .....	15
Figura 2 Sistema auditivo. ....	18
Figura 3 Oído Externo. ....	20
Figura 4 Partes del sistema auditivo. ....	22
Figura 5 La cóclea y el rango de frecuencias.....	24
Figura 6 Plano tridimensional para la locación de fuentes sonoras.....	26
Figura 7 Diferencia de Tiempo Interaural (ITD) y Diferencia de Intensidad Interaural (ILD).....	27
Figura 8 Cono de confusión.....	28
Figura 9 A la izquierda el sistema HRIR-oyente. A la derecha una representación de audio espacial con altavoces.....	31
Figura 10 Cámara anecoica para medición de HRTF y esfera obtenida con las respuestas al impulso .....	33
Figura 11 Parámetros básicos de la HRFT. ....	34
Figura 12 Flujo de la señal de un sistema de reproducción binaural. ....	35
Figura 13 Primeros dispositivos de Realidad Virtual de la empresa VPL.....	37
Figura 14 Gafas Oculus Quest 2 .....	41
Figura 15 Gafas Oculus Quest en Home Studio. ....	42
Figura 16 Arquitectura general de un motor de juegos. ....	43
Figura 17 Términos Clave en Unity 3D.....	46
Figura 18 Representación del Ambisonic y el orden de los armónicos.....	49
Figura 19 Paul McCartney con la consola TG12345.....	54
Figura 20 Interfaz gráfica de Pro Tools .....	56
Figura 21 Esquema Básico de los modelos de procesos.....	59
Figura 22 Modelo de Prototipo.....	60
Figura 23 Ciclo de vida del modelo de prototipo .....	61
Figura 24 Rayo Teleport .....	66
Figura 25 Diseño del menú de mano .....	66

Figura 26 Audio Mixer de Unity .....	68
Figura 27 Prototipo inicial a la izquierda y a la derecha el modelo de escritorio con las bases .....	69
Figura 28 Cuarto de control, escenario principal.....	73
Figura 29 Cuarto de control.....	75
Figura 30 Simulación de acondicionamiento acústico. ....	76
Figura 31 Cuarto de Mezcla.....	77
Figura 32 Representación de los Sonidos .....	78
Figura 33 Entorno Aula .....	79
Figura 34 Menú de mano .....	80
Figura 35 Menús móviles, a la izquierda la barra de transporte, a la derecha el teclado virtual y en el inferior, las guías de usuario.....	81
Figura 36 Puntos de movimiento en el Sistema de Navegación.....	82
Figura 37 Interfaz del Mezclador.....	83
Figura 38 Menú Reproductor de grabaciones.....	84
Figura 39 Reproducción de Videos .....	85
Figura 40 Simulación de sonido espacial .....	86
Figura 41 Logotipo .....	94
Figura 42 tablón con las guías de usuario parte 1 .....	95
Figura 43 tablón con las guías de usuario parte 2.....	96

**Lista de Tablas**

Tabla 1 Listado de Assets y Paquetes .....	99
--	----

## Abreviaturas.

PA → Pabellón Auditivo.

CAE → Canal Auditivo Externo.

CC → Células Ciliadas.

CCI → Células Ciliadas Internas.

CCE → Células Ciliadas Externas.

DAW → Digital Audio Workstation.

HRIR → Head Related Impulse Response.

HRTF → Head Related Transfer Function.

XR → Realidad Extendida.

RV → Realidad Virtual.

VA → Virtualidad aumentada.

RA → Realidad aumentada.

RM → Realidad mixta.

OOP → Programación orientada a objetos.

SDK → Software Development Kit.

API → Application Programming Interface.

IED → Integrated Environment Development.

UI → Interfaz de Usuario.

# 1 Introducción

## 1.1 Problema de investigación

La producción de audio gira en torno al uso de herramientas digitales que permiten la creación y captura de sonidos, su edición, organización, procesamiento y aplicación de efectos. Estas herramientas varían según el tipo de contenido y los dispositivos de reproducción. De manera tal, que existe software especializado para producción musical, así también como para cine, televisión y videojuegos. Actualmente, existe una gran variedad de programas informáticos enfocados para la grabación de instrumentos y voces, para la edición, mezcla y masterización de canciones, composición o diseño sonoro. A estos programas se les conoce como Estaciones de Trabajo de Audio Digital (Digital Audio Workstation o más conocido como DAW) y son los encargados de ordenar y procesar los audios en pistas individuales. Por lo general, estos DAW generan contenido para distintos sistemas de reproducción, en el caso de la música trabajan para sistemas estéreo, mientras que en el caso del cine, tv y videojuegos trabajan con sonido envolvente o sonido espacial, conocido también como sonido 3D.

El sonido espacial es fundamental para crear experiencias inmersivas, puesto que brinda al usuario la sensación de estar dentro de la película o el videojuego mediante la percepción auditiva, simulando los efectos acústicos de los sonidos en relación con el espacio y la anatomía del oyente. En ese sentido, las aplicaciones en Realidad Virtual aprovechan al máximo el audio espacial para lograr mejores experiencias, adquiriendo mayor realismo. Esto se hace mediante algoritmos con cálculos matemáticos avanzados como las Funciones de Transferencia Relacionadas con la Cabeza (HRTF) y promedios estadísticos de las características fisionómicas de los seres humanos, en cuanto a formas del oído, hombros y cabeza. El sonido espacial tiene varias clasificaciones de acuerdo con el sistema de reproducción para el cual ha sido diseñado. Uno de estos es el audio 3D binaural, y utiliza audífonos para reproducir sonidos provenientes de cualquier lugar estando dentro de un ambiente virtual. Por ejemplo, permite que los sonidos ubicados detrás de un usuario sean percibidos realmente como si vinieran de atrás y no de los lados, de adelante o de ninguna

parte, generando una mejor inmersión ya que permite escuchar los sonidos con una ubicación espacial aproximada.

De acuerdo con lo expuesto, últimamente han surgido tendencias para producir música con efectos de espacialidad, y aunque existen subprogramas para los DAW los cuales generan estos efectos en los sonidos, no existía una herramienta en Realidad Virtual que permitiera tener una experiencia inmersiva manipulando audios espacialmente. Por tanto, se propuso el desarrollo de una aplicación en Realidad Virtual que permitiera la manipulación de pistas de audio individuales, la cual tuviera capacidad para mezclar audio tridimensional y que permitiera reproducir en audífonos las grabaciones que se realizaran con esas pistas de audio.

La aplicación realizada brinda al usuario una experiencia inmersiva relacionada con producción musical, con estudios de grabación y principalmente con sonido espacial. Para lograr esa experiencia, se diseñaron entornos que permitieron una relación directa con el sonido por medio de representaciones visuales, las cuales se utilizaron como referencias de la ubicación puntual de las fuentes sonoras. Por lo cual, la aplicación tuvo en cuenta parámetros del audio como el volumen, la frecuencia, y la ubicación espacial en cuanto al acimut y elevación de cada sonido. De esta manera, el desarrollo de la aplicación contó con los menús y entornos necesarios para garantizar la inmersión e interacción del usuario dentro de la Realidad Virtual.

En síntesis, este proyecto de investigación aplicada propuso un desarrollo tecnológico que permitió generar una experiencia inmersiva relacionada con sonido espacial, y que por lo tanto, buscó investigar sobre la mejor manera para desarrollar una aplicación en Realidad Virtual que permitiera mezclar pistas de audio espacializadas, la cual contara con la capacidad para obtener un archivo de audio en formato WAV.

## **1.2 Objetivos de la investigación**

### **1.2.1 Objetivo general**

Desarrollar una aplicación en Realidad Virtual que permita mezclar pistas de audio individuales para obtener un archivo en formato WAV, mediante interacciones con sonido espacial.

### **1.2.2 Objetivos específicos**

- Diseñar 3 escenarios virtuales, en donde se simule un estudio de producción musical, un ambiente de mezcla para sonido espacial y un ambiente extra para informar sobre los conceptos utilizados en la aplicación.
- Programar las mecánicas de interacción de los escenarios y de los sonidos con respecto al volumen, la frecuencia y la ubicación espacial.
- Implementar un sistema de grabación que permita almacenar en un archivo formato WAV, los sonidos que se mezclen dentro de la aplicación.

### **1.3 Justificación de la investigación**

Los recientes avances de la tecnología en dispositivos de Realidad Virtual generan líneas de investigación que en un futuro podrían ser nodos de exploración tanto en la Universidad Pedagógica Nacional de manera local, como a nivel nacional e internacional en otras instituciones. En ese sentido, el uso de Realidad Virtual en diversas áreas del conocimiento y la industria ha motivado un crecimiento de nuevas aplicaciones que son transversales a todos los campos del saber. Para este y futuros proyectos, el uso de Realidad Virtual en producción musical podría significar un avance en el tratamiento y comunicación de la información relacionada con audio. Por este motivo se propuso una investigación que tuviera que ver con la creación de nuevas interfaces de usuario enfocadas hacia la mezcla visual, haciendo uso de dispositivos de Realidad Virtual. Al respecto sobre la posibilidad de imaginar nuevas relaciones con el sonido y el audio, existían varios trabajos predecesores que sugirieron la pertinencia y la necesidad de investigar sobre este tema, basado en la implementación de tecnología de punta.

Por otra parte, la creación de aplicaciones en Realidad Virtual puede impactar positivamente en la manera en que se realizan las producciones musicales y la manera en que se enseña y se aprende de estas. Porque de manera interactiva se pueden lograr experiencias significativas que relacionen el espacio con los sonidos. Se pueden imaginar otras formas de practicar y explicar conceptos clave en el mundo de los estudios de grabación y en el procesamiento digital de señales, haciendo uso de nuevos dispositivos. Para que los estudiantes afines a la música, la producción musical, la programación y el desarrollo de

software, tengan mayor variedad de materiales de apoyo para su estudio y entretenimiento. De esta manera, el uso de dispositivos de Realidad Virtual en educación contribuye a nuevas relaciones con el saber y a su vez permite su implementación de diferentes maneras, en este caso, con actividades relacionadas a la generación de contenido audiovisual.

Finalmente, este desarrollo estuvo sujeto a los trabajos que se han venido realizando en el semillero de investigación del grupo Alice de la Licenciatura en Electrónica, en donde se estudian temas relacionados con sonido espacial para su implementación en aplicaciones de Realidad Virtual, enfocadas hacia las ciencias naturales, la simulación de entornos para capacitación y videojuegos educativos, en donde el audio es pieza clave para lograr mejores experiencias inmersivas. Por tanto, este trabajo investigativo permitió una mayor comprensión del sonido espacial aplicado a desarrollos en Realidad Virtual, lo cual fue un avance significativo para el semillero. Puede ser un material de apoyo para estudiantes tanto de la licenciatura, como para productores aficionados y desarrolladores, ya que puede ser implementado en ambientes formativos como también en ambientes de entretenimiento y de simulación.

#### **1.4 Producto final entregable**

Se entrega una aplicación en Realidad Virtual con escenarios interactivos que permiten la obtención de un archivo de audio basado en la mezcla de sonidos individuales espacializados. La aplicación cuenta con los siguientes ambientes:

- 1- En el escenario principal (o cuarto de control), se ingresa a un estudio de grabación con menús de navegación y guías de usuario.
- 2- En el escenario central (o cuarto de mezcla), el usuario está dentro de un recinto cerrado con opciones para mezclar, grabar y reproducir sonido espacial.
- 3- En el tercer escenario, se encuentra información sobre los conceptos de audio y de mezcla utilizados en la aplicación, así también como un tutorial sobre el buen manejo de las interfaces gráficas de usuario.

## 2 Antecedentes y Marco Teórico

### 2.1 Antecedentes

La presente investigación tiene como referente primario las investigaciones y trabajos desarrollados por (Gibson, *The Art Of Mixing A Visual Guide To Recording Engineering And Production.* , 1997), en donde el autor desarrolló un sistema de mezcla basado en la visualización de imágenes relacionadas con las características de los sonidos y los efectos agregados a los mismos dentro de un espacio tridimensional. Según la patente, el sistema involucra una consola de mezcla con canales de entrada individuales, un microcomputador que transforma las señales de audio en imágenes, una pantalla, unos altavoces y el código de programa que permite la manipulación de los sonidos y la visualización de éstos en tiempo real. Esta obra es única dentro de la literatura de ingeniería de sonido, pues expone un modelo pedagógico sobre las herramientas y técnicas tradicionales para lograr una gran mezcla estéreo y propone el uso de informática de audio y Realidad Virtual, en conjunto con consolas de mezcla tradicionales. El concepto aquí desarrollado sobre la visualización de la imagen de los sonidos fue la inspiración para el presente desarrollo y por tanto, uno de los ejes principales.

En el mismo sentido de las ideas de Gibson, (Agostini, 2014) en su trabajo de maestría en ingeniería de sonido, presentó un prototipo que hace posible la mezcla de una canción en un espacio tridimensional con el uso de un sensor de seguimiento de mano llamado Leap Motion. Este documento aporta información teórica importante al detallar la naturaleza del sonido desde la psicofísica, así también como la introducción del concepto de objeto sonoro para tratar los sonidos dentro del motor de desarrollo. En sus disertaciones invita a continuar trabajando en nuevas interfaces naturales de usuario enfocadas en la mezcla visual, en particular la utilización de Realidad Virtual, por lo cual el presente desarrollo es un intento sucesivo en esa misiva.

Por otra parte, el trabajo de (Rodríguez, B.) sugirió el desarrollo de un plugin para la estación de trabajo de audio digital (DAW) Ableton Live 8, el cual permitió representar los conceptos de Gibson en una interfaz pequeña, visualizando los sonidos de cada canal en el secuenciador. En ese mismo campo, empresas dedicadas al desarrollo de plugins para audio,

también han creado entornos similares al anterior, como es el caso de la empresa Izotope, quienes han lanzado al mercado un plugin llamado Visual Mixer, el cual permite enlazar canales individuales del DAW a un canal máster para que los sonidos sean visualizados mediante esferas y éstas puedan ser localizadas en ubicaciones virtuales-espaciales.

Otro referente importante, es el trabajo que ha realizado (Bes, 2020), quien ha desarrollado una emulación en Realidad Virtual de una consola analógica AMS - NEVE VR88. Esta consola tiene una trascendencia especial siendo un hardware legendario, y el autor realizó una aplicación que permite ingresar en un estudio de grabación e interactuar con la consola, controlando todos los mandos del tablero de control y enviarlos al hardware físicamente. Es de especial relevancia este trabajo puesto que el desarrollador utilizó el motor de Unity para su elaboración junto con Blender. De esta investigación no se cuenta con un documento formal, pero existe información en la página oficial de la empresa Anymotion formada por el mismo autor. Adicionalmente, se tuvo contacto directo con Bes quién compartió conocimientos sobre el desarrollo de su aplicación.

Adicionalmente, un antecedente relevante es el artículo elaborado por (Gelineck & Overholt, 2016) “Haptic and Visual feedback in 3D Audio Mixing Interfaces”, donde se realizó una comparativa entre la realimentación háptica producida en una interfaz de mezcla en 3D, usando dos tipos de controladores, un Leap Motion y un Novit. En su investigación evaluaron la interfaz con 3 personas quienes afirmaron una mejor percepción sensorial con el Novint, puesto que les permitía una realimentación háptica (del tacto) sin tener que mirar a la pantalla. Mientras, con el Leap Motion, la sensación no fue satisfactoria puesto que esta tecnología no ofrece una realimentación sensorial sin dejar de mirar el monitor. Esta investigación es relevante porque proporciona información acerca de la reacción de las personas ante controladores que no brindan una realimentación sensorial mediante el tacto.

En el trabajo de grado de (Gaitán Quiasua, 2015) “Implementación del Leap Motion como controlador midi interactivo mediante reconocimiento gestual (dactilar)”, la autora desarrolló un controlador midi no convencional basado en el reconocimiento de los gestos de la mano mediante tecnología Leap Motion. Allí, realizó una interfaz en Unity 3D que le permitió controlar via midi una estación de audio digital (DAW) como Ableton Live. Hizo el análisis pertinente para el reconocimiento gestual de las manos y así logró controlar diferentes

parámetros del software de producción. Lo importante en esta investigación es la comunicación entre el sensor, el DAW y Unity 3D, así como los procesos de comunicación mediante Pure Data y protocolos MIDI, OSC-UDP.

También es interesante destacar el trabajo de (Carrascal & Jordà, 2011) quienes desarrollaron un prototipo de interfaz multitáctil para mezcla de audio, en donde se enfocaron inicialmente en una interfaz de usuario para sonido envolvente, y en obtener una mecánica de interacción distinta al diseño basado en potenciómetros de las consolas de mezcla tradicionales. La aplicación fue hecha para Reactable haciendo uso de protocolos de comunicación OSC junto con Pure Data. Tuvo una etapa de análisis y evaluación por 6 personas obteniendo tiempos de mezcla distintos al hacer uso de la interfaz virtual y de un mixer tradicional. Los resultados fueron positivos para la aplicación y mostraron la importancia de nuevas interfaces de usuario para manejo inexperto y la necesidad de agregar funciones más complejas para usuarios expertos.

En otro estudio realizado por (Brinkman, Hoekstra, & Egmond, 2015) se muestran los resultados de tres pruebas realizadas a una serie de personas entre mujeres y hombres de 27 y 28 años, en donde, primero tenían que distinguir entre cuatro técnicas de audio: mono, estéreo, dolby surround y audio 3D. Segundo, tenían que indicar el grado de presencia, ansiedad y percepción espacial de las cuatro técnicas. Y tercero, los participantes experimentaban las cuatro técnicas de audio en conjunto con estímulos visuales. Todas las pruebas se hicieron con el sonido de una avispa. En el primer lugar, el sonido por sí mismo puede provocar ansiedad. En segundo lugar, si sólo se proporciona un estímulo de audio, la experiencia de las personas se ve afectada por el tipo de técnica de audio. Tercero, añadir sonido a un entorno visual puede mejorar la experiencia, pero parece poco probable que comparado con el sonido estéreo, el audio 3D no añada algo a la experiencia de las personas cuando se exponen a un ambiente de estímulos sólo con audio, o a un mundo de audio combinado con estímulos visuales. Lo interesante de este trabajo es que permitió afirmar que el audio 3D sí tiene un impacto en la percepción de lo visual, en este caso, aplicado a Realidad Virtual, aunque sea posible que una persona no calificada no logre distinguirlo del sonido estéreo.

El trabajo de grado de (Cardona Rojas & Valderrama Pulgarín, 2014) acerca del desarrollo de un plugin VST para la reproducción de audio binaural por medio de bases de datos de Funciones de Transferencia Relacionadas con la Cabeza (HRTF), es un documento clave para el entendimiento tanto del desarrollo de plugins, como para la comprensión matemática de las HRTF. Pues en este trabajo se expuso con claridad, el carácter de filtro digital que se requiere implementar para reproducir en tiempo real la posición de una fuente sonora en un entorno virtual. En ese mismo sentido, un trabajo anterior hecho por (Barrera Giraldo, 2009) desarrolló un prototipo de espacializador de sonidos monofónicos en una aplicación para Windows. En este trabajo se expone en amplitud la historia y teoría alrededor del sonido espacial. Se detalla con precisión la fisionomía del sistema auditivo y la forma en la cual el ser humano logra transducir los sonidos externos. En los dos anteriores documentos se realizaron pruebas subjetivas a un grupo de personas para determinar qué tan preciso se podían localizar las fuentes sonoras en espacios virtuales. Y en ambos casos, se tuvo mayor rango de error en la ubicación vertical, debido a las bases de datos genéricas de las HRTF. Pues se debe tener en cuenta que utilizar HRTFs de forma personalizada, es lo que posibilita una percepción propia de escucha.

Desde el contexto planteado por los documentos, artículos, libros, aplicaciones, tesis e investigaciones anteriormente descritos, se sustenta el presente desarrollo. Puesto que aportaron información crucial para definir el rumbo del proyecto, en cuanto a los aspectos que en relación con los objetivos debería tener la aplicación, la metodología, el diseño, las mecánicas y las interfaces de usuario.

## **2.2 Marco Teórico - Conceptual**

A continuación, se examina la teoría que hace posible trabajar con audio espacial y con Realidad Virtual. Se explora en los estudios sobre la localización espacial sonora y se describe el sistema auditivo humano. Se plantea el mecanismo auditivo, como un complejo sistema que permite el entendimiento razonable de eventos sonoros externos, para de este modo comprender el modelo matemático que posibilita la espacialización de los sonidos mediante funciones de transferencia. Se expone información sobre Realidad Virtual, su

historia, dispositivos recientes, mejores prácticas y técnicas, para finalizar con los motores de desarrollo y las herramientas tradicionales en los estudios de producción musical.

El marco teórico relaciona la teoría científica con el problema investigativo, para de esta manera crear una referencia que sitúa el problema en un punto susceptible de ser contrastado. Pues al ser el marco teórico la base de respaldo de una investigación, éste puede reformularla y hasta cambiar sus objetivos (Tamayo y Tamayo, 2003). Por tanto, el mayor esfuerzo en la creación de este proyecto se encuentra a continuación.

## **2.2.1 Sonido y Señales**

### **2.2.1.1 Naturaleza del sonido**

Existen distintos tipos de ondas y movimientos. En física, se le llama movimiento armónico simple a los cambios periódicos de posición que implican una fuerza perturbadora y una fuerza restauradora, que siempre conlleva finalmente a un punto de equilibrio. Este tipo de movimiento es análogo a lo que sucede con un resorte, cuyos extremos están referenciados a un punto fijo y a un objeto con masa en movimiento. El resorte posee una fuerza que siempre va a retornar al objeto a una posición de reposo<sup>1</sup> (siempre y cuando el resorte no pierda su forma y elasticidad). Por tanto, los conceptos del movimiento armónico simple son funcionales para todo objeto cuya vibración se propaga a través de un medio, como las ondas mecánicas.

Existen diferentes tipos de ondas y se clasifican de acuerdo con la manera en que se propaga su vibración y la fuente que las produce. Por ejemplo, hay ondas de agua, ondas sísmicas, ondas de radio y las que son de interés de este estudio, ondas acústicas. El ejemplo más utilizado para explicar qué es una onda, se remite a pensar en una roca arrojada a un lugar con agua. Al caer la roca se genera una perturbación en la superficie del agua creando círculos que se desplazan hacia afuera del punto de impacto de la roca, moviéndose horizontal y verticalmente hacia arriba y abajo. Esto sugiere que para que exista una onda debe haber primero que todo, una fuente que genere una perturbación, un medio por el cual se pueda propagar el movimiento de esa perturbación y un mecanismo que relacione la influencia entre partículas alrededor de la perturbación (Serway, Vuille, & Faughn, 2009). De este

---

<sup>1</sup> A esta acción del resorte se le conoce como la ley de Hook y fue demostrada por Roberto Hook.

planteamiento se infiere que una onda debe contener energía cambiante de un objeto en vibración, y que esa energía debe generar movimiento de moléculas y electrones en un medio (McCarthy, 2007).

Existen dos tipos de ondas, las cuales difieren según el tipo de desplazamiento que genera la perturbación. Las ondas que se desplazan perpendicularmente a lo largo del medio se conocen como transversales, y un ejemplo se da cuando una cuerda sujeta a un punto fijo es movida periódicamente en el eje vertical desde el extremo libre, haciendo que en la cuerda se creen una especie de pulsos. Estos pulsos recorren la cuerda horizontalmente mientras los movimientos que generan los pulsos son efectuados verticalmente. El ejemplo anterior del agua se trata en cierta forma de una onda transversal, pero aún hay más. Por otra parte, las ondas que se mueven en paralelo con la dirección de propagación, se les conocen como ondas longitudinales. Y como ejemplo, es posible pensar en un resorte al cual se le aplica una fuerza periódica en la misma dirección en la que el resorte se contrae y se estira. Las partes del resorte contraídas se pueden pintar en un gráfico de densidad versus tiempo, en donde las zonas con mayor contracción tendrán mayor densidad y viceversa, pintando una forma sinusoidal. Este gráfico de crestas y valles representa también, la rarefacción y compresión del aire producido por el sonido, en donde las ondas mecánicas generan zonas de mayor y menor presión sonora (Agostini, 2014).

En adelante a los sonidos se les denominará fuentes sonoras, las cuales pueden ser ocasionadas por cualquier objeto que entre en vibración<sup>2</sup>. Esa vibración hace que las partículas del medio oscilen alrededor de su punto de equilibrio generando que las demás partículas también oscilen. Una vez la fuente sonora se detiene el sonido se acaba, y las partículas que se estaban moviendo adelante y atrás, nunca se desplazaron hacia ninguna parte, pues siempre volvieron a su origen, igual que el resorte (McCarthy, 2007). Esto provoca un movimiento en el medio el cual es llamado como propagación de onda sonora (ver figura 1), y la forma de medir el cambio total que sufre el medio es a través de la presión sonora ( $P_t$ ), la cual tiene como referencia la presión atmosférica normal ( $P_0$ ), más los

---

<sup>2</sup> Para que puedan ser percibidas deben estar dentro de las ondas audibles, que son ondas longitudinales de 20Hz a 20.000Hz que es el rango de audición humano. Ese rango frecuencial es difícil de obtener en la mayoría de altavoces, por lo cual se podría acotar de 31Hz a 18KHz. También hay ondas infrasónicas y ultrasónicas las cuales están por debajo o por encima de la percepción del oído humano.

incrementos y disminuciones debidos al comportamiento de la fuente sonora denominados como (P), los cuales varían positiva y negativamente en el tiempo.

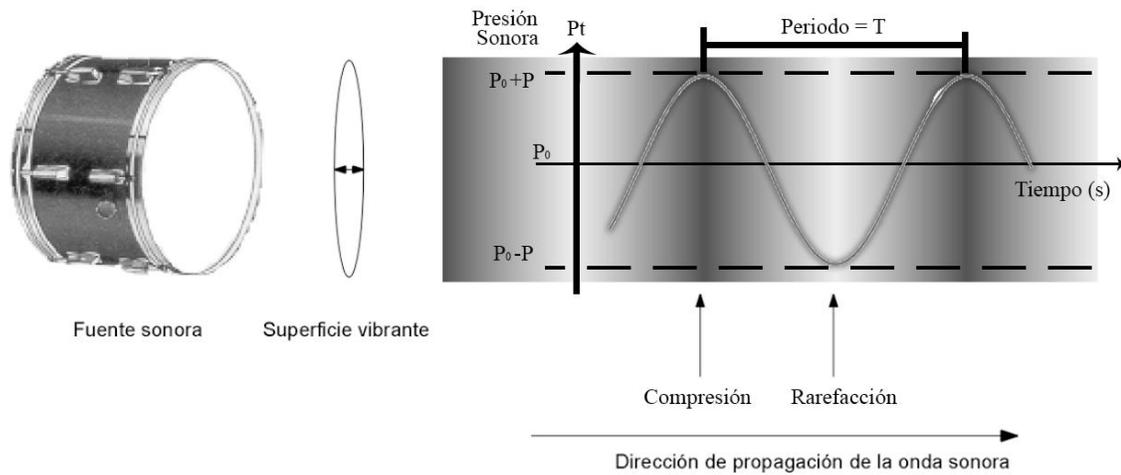


Figura 1 Forma sinusoidal de las zonas de rarefacción y compresión de las partículas del aire

Fuente: modificado de (Carrión Isbert, 1998)

En la figura 1 se puede observar un ejemplo propuesto por (Carrión Isbert, 1998) en donde un tambor es la fuente de sonido que genera una perturbación. Allí se puede observar la forma sinusoidal que adopta la propagación por el medio, en ese caso, el aire. Se puede observar que las partículas se comprimen y se enrarecen cambiando de intensidad a lo largo del tiempo. A ese cambio en los valores de intensidad se le conoce como amplitud de onda o magnitud, y es una medida referente al cambio máximo y mínimo que produce la fuente en el medio, o también del nivel de presión sonora que ejerce sobre el medio. Al tiempo que toma de ir desde un punto de igual magnitud a otro, o también, el tiempo que tarda la señal en completar una oscilación, se le denomina Periodo y en la mayoría de los casos se da en milisegundos (ms). El inverso del tiempo es lo que se conoce como la frecuencia, la cual se mide en Hertz (Hz) e indica la cantidad de ciclos por segundo de una señal. Para poder ubicarse en algún punto de la señal sin importar la posición en el tiempo se utiliza el concepto de fase, el cual se da en grados y puede ir de 0 a 360. Otro parámetro importante es la longitud de onda, cuyos valores se dan en metros o pies. La longitud de onda es la distancia física, real, que existe entre un ciclo y otro en un medio específico. La expresión matemática de la longitud de onda indica que su tamaño para una frecuencia dada depende de la velocidad de propagación del medio. Y la velocidad de propagación depende de la temperatura del medio,

por tanto, en un ambiente con temperaturas cambiantes la longitud de onda, igual que la velocidad de propagación será cambiante. Si la temperatura aumenta, la longitud de onda se expande y también la velocidad de propagación, puesto que las partículas del aire tendrán mayor energía. Se conocen cambios de aproximadamente 1% en la velocidad de propagación por cada 5°C de aumento o decremento en la temperatura (McCarthy, 2007).

Expresiones matemáticas:

Periodo y Frecuencia

$$T = \frac{1}{F}; \quad F = \frac{1}{T}$$

Longitud de Onda

$$L = \frac{c}{F}$$

Donde “c” es la velocidad de propagación del sonido en el aire, la cual depende de la temperatura en grados centígrados. 331.4 es la velocidad del sonido a 0 grados centígrados.

$$c = 331.4 \frac{m}{s} * \sqrt{\frac{T}{273K}} \quad \text{sí T es } 22^{\circ}\text{C, en Kelvin serian } 295K$$

$$c = 344.49 \text{ metros/segundo}$$

La forma de onda sinusoidal descrita con la propagación de onda sonora se puede expresar matemáticamente mediante la ecuación de movimiento armónico. Donde A es la amplitud, omega es la frecuencia angular ( $\omega = 2 * \pi * f$ ) y gamma es la fase en radianes<sup>3</sup>.

$$f(t) = A * \sin(\omega * t + \gamma)$$

Por otra parte, la relación entre la potencia acústica y la superficie por la cual el sonido atraviesa se conoce como intensidad acústica<sup>4</sup> y está dada en  $watts/metro^2$ . Sobre esta base se obtienen los rangos de audibilidad, los cuales están referenciados a una frecuencia de

---

<sup>3</sup> Tomado de (Agostini, 2014).

<sup>4</sup> También se conoce como intensidad de onda.

1000Hz<sup>5</sup>. Para el umbral de audibilidad, que es el sonido más bajo que el oído puede percibir, se tiene una intensidad acústica de  $1 \times 10^{-12} \text{W/m}^2$ . Y para el umbral de dolor, que es el sonido más intenso que el oído puede soportar se tienen intensidades del orden de  $1 \text{W/m}^2$ . El oído es un detector de ondas sonoras de gran precisión, y la forma en la que percibe la intensidad de los sonidos no es lineal y tampoco sería práctico un manejo lineal de los niveles de intensidad (Serway, Vuille, & Faughn, 2009). De lo anterior, la diferencia entre umbrales es de  $10^{12}$ , lo cual sugiere que el oído percibe la intensidad sonora de una manera logarítmica, puesto que un aumento entre el sonido más débil y el sonido más fuerte definitivamente no se percibe como de  $10^{12}$ . Por otra parte, los sonidos no son estáticos, la presión varía con el tiempo. Por tanto, para tener una medición aproximada del valor de presión acústica, se obtiene el promedio temporal o el valor eficaz (RMS). La presión se da en Pascales (Pa) y el oído también tiene un rango para soportar la presión acústica. En el umbral de audición el oído es capaz de percibir presiones entre  $2$  y  $3 \times 10^{-5} \text{Pa}$  y en el umbral de dolor hasta  $100 \text{Pa}$ , sobre una presión atmosférica de  $1 \times 10^5 \text{Pa}$ . Esto último, reconfirma que para percibir un rango de presiones de  $1$  a  $5$  millones, el oído debe tener un comportamiento logarítmico. Por tanto, el uso del decibel<sup>7</sup> (dB) es imperativo para lograr simplificar los análisis. Y para referirse al nivel de intensidad<sup>8</sup> ( $\beta$ ) se debe tener la intensidad a convertir ( $I$ ) y el umbral audible ( $I_{ref}$ ) en la siguiente expresión matemática.

$$\beta = 10 * \log\left(\frac{I}{I_{ref}}\right) = 10 * \log\left(\frac{I}{1 \times 10^{-12} \text{W/m}^2}\right)$$

El uso de un valor de referencia como el umbral audible tiene su propósito en el permitir que todos los valores de nivel sean positivos. El nivel de presión sonora (SPL) es una medida estándar que permite saber que tan intenso es un sonido mediante una referencia que es el umbral audible, esto permite que se pueda trabajar un rango entre  $0$  y  $135 \text{dB}$ <sup>9</sup>, siendo el  $0$  el nivel más bajo que se puede escuchar (no implica ausencia de sonidos) y  $135$  el umbral de dolor (Carrión Isbert, 1998). En el nivel de  $0 \text{dB SPL}$  se tiene una presión de  $0.0002$

---

<sup>5</sup> Se toma como referencia los  $1000 \text{Hz}$  porque es la zona frecuencial con mayor distancia entre umbrales (Berenguer Alarcón, i Castellanes, & Jordà Puig, 2005).

<sup>6</sup> La relación de sensibilidad del oído es de una en un trillón.

<sup>7</sup> El decibel es un sistema de escala logarítmica que relaciona rangos de valores muy grandes o pequeños.

<sup>8</sup> Cabe indicar que el nivel de intensidad es distinto a la intensidad sonora, el primero se da en una escala logarítmica y el segundo en una escala lineal.

<sup>9</sup> Esta diferencia de umbrales se puede analizar como el rango dinámico del sistema aural del ser humano.

$\text{dinas/cm}^2$  el cual se acerca al límite de ruido del aire, producido por el movimiento aleatorio de las moléculas que crean su propio ruido. En la siguiente expresión, P es la presión sonora en valor eficaz.

$$\text{Nivel (db SPL)} = 20 * \log \left( \frac{P}{0.0002} \right)$$

### 2.2.1.2 El sistema auditivo humano

Junto con la visión, el sistema auditivo es uno de los sentidos más evolucionados del ser humano y uno de los más complejos en los vertebrados. Durante siglos, ha evolucionado para que sea posible la supervivencia y la comprensión del entorno. El sistema auditivo se compone de dos orejas, las cuales están preparadas con una gran capacidad para transmitir las señales desde donde vienen los distintos sonidos, y ha sido así porque gracias a estos dos receptores hace miles de años el hombre puede evadir animales al acecho y salir con vida, puede identificar cuando está en peligro y la localización de la fuente sonora que provoca esa sensación. Por ejemplo, al identificar cuando una mosca o un sancudo se encuentra rondando la cabeza o el sonido de un automóvil acercándose. Permite entender la música y provocar cambios de estado anímicos de acuerdo con el tipo de sonidos y la interpretación del lenguaje hablado, involucrando no solamente lo fisiológico y mecánico sino también el sistema nervioso y fundamentalmente el cerebro (García Ruiz, 2020). La función principal del oído es la de amplificar, transformar y filtrar ondas mecánicas producidas por una fuente mediante un medio hacia el tímpano, el cual convierte ese patrón de movimiento ondular en un potencial de acción neural (Soto, Vega, Chávez, & Ortega, 2003).

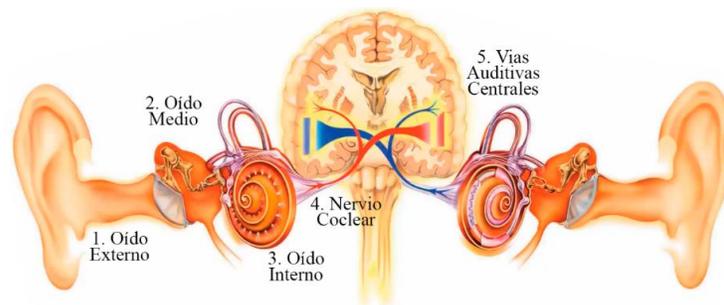


Figura 2 Sistema auditivo.

Fuente: tomado de la ponencia de Marcos Simón en el 1er simposio sobre parlantes y sus aplicaciones.<sup>10</sup>

<sup>10</sup> Información del evento en: <https://www.acusticauach.cl/?p=14022>

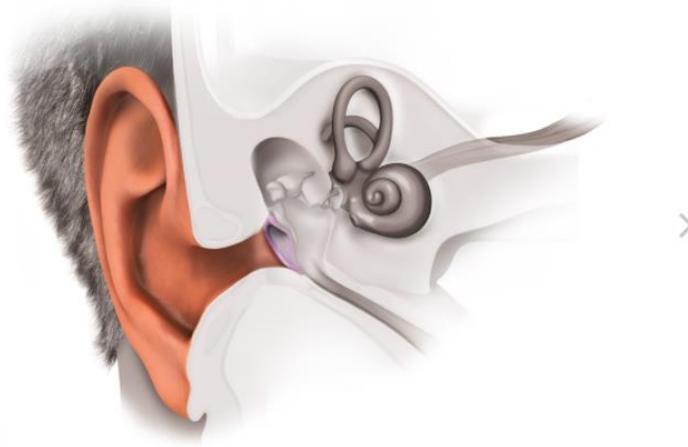
La figura 2 muestra las partes del sistema auditivo, las cuales en (Garcia Ruiz, 2020) se dividen en 5 etapas sucesivas siguiendo el orden fisiológico. En lo siguiente se resumen las primeras 3 etapas en sincronía con varios trabajos referenciados en el presente documento.

- **Etapa 1: Oído Externo.**

Está compuesto por la oreja o Pabellón Auricular (PA), el cual capta los sonidos<sup>11</sup> para enviarlos al Canal Auditivo Externo (CAE) que es la continuación del pabellón y el encargado de convertir las ondas esféricas en ondas planas (Barrera Giraldo, 2009). El PA cumple una función protectora, se encarga de proteger al CAE del viento, dividiendo el aire en distintas direcciones. Sin no se tuviera el pabellón, el golpe del aire evitaría percibir otros sonidos (Berenguer Alarcón, i Castellanes, & Jordà Puig, 2005). Por otra parte, el CAE puede medir entre 2 y 2.5 cm y se encarga de mantener la misma temperatura en el canal y en el tímpano. Es aquí en donde se produce cerumen, más conocido como cera. Esta secreción se genera por higiene y protección ante sonidos intensos evitando que las partículas de polvo ingresen al oído medio y que impactos muy fuertes dañen el tímpano (Miyara, 1999). En esta parte del oído externo se producen efectos interesantes. Inicialmente la forma del canal favorece la percepción espacial de los sonidos, amortiguando las frecuencias bajas permitiendo con facilidad el paso de las frecuencias altas hacia el tímpano según la orientación de la fuente. Análogamente, es como un filtro digital único para cada oreja, que también es llamado como la huella acústica, haciendo alusión a la huella dactilar única en cada ser humano. Finalmente, la forma del canal es como una cámara de resonancia que aumenta la presión acústica de acuerdo con la posición de la fuente, llegando a amplificar hasta 20db en 3000 Hertz y aunque los aumentos se producen en todas las frecuencias, son más notorios en sus múltiplos entre 2 y 7KHZ (Garcia Ruiz, 2020). Esto incide definitivamente en la respuesta frecuencial de cada persona. Más adelante se detalla en qué medida esto afecta la percepción espacial de los sonidos.

---

<sup>11</sup> Desde un punto de vista físico, todo cambio de presión en el movimiento de las partículas que se propagan en un medio elástico se comprende como sonido. El sonido también es susceptible de ser transmitido por medios distintos al aire desde que el medio sea elástico, tenga masa e inercia. Por tanto, es posible que se transmita en sólidos y en líquidos. Incluso la velocidad del sonido puede aumentar en un medio sólido debido a que las partículas se encuentran más juntas, evitando pérdida de energía por desplazamiento al oscilar alrededor de su punto de equilibrio. La temperatura ambiente es otro factor que hace que aumente la velocidad del sonido en la propagación por un medio como el aire. Esto se detalla más adelante.



*Figura 3 Oído Externo.*

*Fuente: tomado de med-el. How hearing works<sup>12</sup>*

- **Etapa 2: Oído Medio.**

Está formado por el tímpano, los huesecillos y la trompa de Eustaquio, los cuales se encuentran dentro de la caja timpánica. En esta parte del oído también hay músculos y membranas. Especialmente se realiza una conversión del medio acústico que viene del CAE (aire) al medio sólido de los huesecillos y al líquido linfático del oído interno. Por este motivo se dice que el oído medio funciona como un acople de impedancias acústicas. Durante el paso del sonido por las primeras dos etapas, este va perdiendo energía debido a la resistencia o impedancia del medio, la cual está dada por la densidad y la elasticidad de este. La velocidad del sonido en un medio como el aire a una presión atmosférica normal viaja a 340 m/s, variando 1 grado cada 5.5 grados Celsius (Barrera Giraldo, 2009) y en un sólido como el hierro viaja a 3000 m/s. Cuando una onda sonora intenta cambiar de medio acústico es posible que pierda o gane energía. Por ejemplo, el paso de un medio como el aire a uno como el agua, implica la reflexión de hasta el 99% de energía sonora. En el oído medio, esta transformación de impedancias se lleva a cabo mediante dos mecanismos, uno de acción hidráulica en donde la diferencia de superficies entre el tímpano, el estribo y la ventana oval supone una amplificación media de 22 veces la presión acústica y otro mediante la acción de las palancas de primer grado formada por la cadena de huesecillos. A este respecto cabe indicar que el tímpano no vibra igual en todas las frecuencias, este tiene movimientos

---

<sup>12</sup> Obtenido de: <https://www.medel.com/about-hearing/how-hearing-works>

específicos de acuerdo con la frecuencia e intensidad de la fuente sonora, por ello la vibración de este depende de las características de cada sonido. Los más recientes estudios indican frecuencias de resonancia del tímpano entre 1Khz y 2Khz, donde las frecuencias bajas hacen que la membrana timpánica vibre en el centro y las frecuencias mayores a 2Khz hacen que vibre por completo (García Ruiz, 2020). Por otra parte, no todas las vibraciones del tímpano provocan un movimiento en la cadena de huesecillos, algunas vibraciones se transmiten al oído interno via gaseosa, estimulando el aire de la caja timpánica hasta llegar a la ventana redonda (Berenguer Alarcón, i Castellanes, & Jordà Puig, 2005).

Los huesecillos funcionan como un conjunto de palancas con una relación de presión mayor a la que funciona el oído externo, aquí se convierten las vibraciones de gran amplitud y poca presión del tímpano, a movimientos de poca amplitud (nanómetros) pero de mayor presión ejercidos por la platina del estribo hacia la ventana oval, la cual funciona como la puerta de comunicación del oído medio al interno. Este cambio de fuerzas es lo que en un evento sonoro súbito como una explosión puede hacer que el oído se dañe, puesto que la fuerza con la que el estribo impacta a la ventana oval puede destrozarla. Para evitar esto, y cuando los sonidos no tienen un tiempo más corto que la respuesta de los huesecillos (500ms), los músculos del tímpano y el estribo se contraen limitando el movimiento de la cadena, evitando que un exceso de presión llegue al oído interno mediante el impacto del estribo.

En síntesis, las ondas mecánicas en vibración provenientes del CAE ingresan hasta el oído medio y generan movimiento ondulatorio en el tímpano el cual transmite esa energía a los huesecillos (martillo, yunque y estribo), en donde el tímpano es el puente entre el canal auditivo y la caja timpánica. Para que el tímpano pueda vibrar en sincronía con las variaciones de presión acústica, la función de la trompa de Eustaquio debe ser la de comunicar la caja timpánica con la faringe para igualar la presión de la caja con la presión atmosférica, esto para que el oído medio funcione independientemente de las condiciones atmosféricas (Berenguer Alarcón, i Castellanes, & Jordà Puig, 2005) y para que se mantenga una presión constante (Miyara, 1999). Los cambios de presión súbitos provocan una sensación de abombamiento en la cabeza (o de oídos tapados), como cuando se sube a una montaña, o se profundiza en el agua, o se sale de la ciudad hacia un lugar con menor nivel

sobre el mar. Este cambio genera un corte en las frecuencias altas hasta que la trompa de Eustaquio equilibre la presión interior del oído y esto se hace inconscientemente al momento de bostezar o tragar saliva. A este respecto, cabe indicar que un cambio brusco de presión en el ambiente puede llegar a deformar el tímpano hasta el punto de daño irreversible. Para una mejor comprensión del proceso descrito hasta aquí, ver la figura 4.

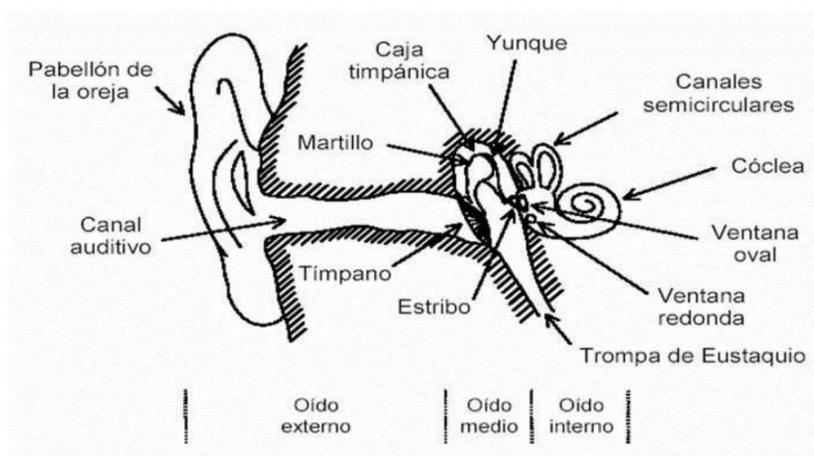


Figura 4 Partes del sistema auditivo.

Fuente: tomado de Castrillo, C. B. Capítulo I. Psicoacústica<sup>13</sup>

- **Etapas 3: Oído interno.**

Conformado principalmente por la cóclea, por la ventana oval, la ventana redonda y los conductos semicirculares, el oído interno se encuentra ubicado en una cavidad del hueso temporal, uno de los huesos más duros del cráneo. Cuando las vibraciones provenientes del oído medio llegan a la ventana oval y a la ventana redonda, consiguen alcanzar la cóclea y empujar células especializadas en convertir las señales acústicas en impulsos nerviosos eléctricos. Estas células son llamadas Células Ciliadas (CC) y su funcionamiento se asemeja al de unos pequeños micrófonos que transducen la variación de presión acústica en señales bioeléctricas<sup>14</sup> del orden de los 80 a 90mV. Estas señales se conocen como potenciales de acción y son capaces de ser interpretadas por el Sistema Nervioso Central (SNC). Esa diferencia de potencial aproximada se genera porque dentro de la cóclea, existen canales

<sup>13</sup> Obtenido de: <https://bibing.us.es/proyectos/abreproy/70661/fichero/Cap%C3%ADtulo+I.+Psicoac%C3%BAstica..pdf>

<sup>14</sup> Esta palabra es acuñada por (García Ruiz, 2020) para referirse a los impulsos nerviosos emitidos por el oído central hacia el cerebro.

conocidos como canal timpánico, canal vestibular y canal coclear. Los cuales tienen dos sustancias llamadas perilinfa, la cual es rica en sodio y endolinfa, rica en potasio. Estas tienen propiedades iónicas contrarias. En el canal timpánico y en el vestibular hay perilinfa y en el coclear hay líquido endolinfático, por ello la diferencia iónica que se genera produce el potencial de acción (Barrera Giraldo, 2009).

La cóclea es un órgano único. Tiene forma de espiral y mide aproximadamente 3,2cm. Podría ser la parte más importante en el proceso de la escucha puesto que es allí en donde se discriminan los sonidos según su componente frecuencial e intensidad en el tiempo. Esta información se codifica y filtra para ser transmitida a las vías auditivas para que posteriormente sean procesadas en la corteza cerebral. En esta parte se han encontrado al menos tres procesos concernientes a dicha discriminación. Primero, lo referente a la mecánica de los movimientos hidráulicos ejercidos por el estribo en la ventana oval y redonda. Segundo, el desplazamiento del órgano de Corti<sup>15</sup> con respecto a la membrana basilar y tercero, la transducción de la energía mecánica en bioeléctrica. Las CC hacen parte del órgano de Corti y este se encuentra en la membrana basilar, la cual está alojada en la mitad entre el canal vestibular y el canal timpánico. La membrana basilar tiene una forma interesante, pues sus dimensiones cambian de acuerdo con la longitud de esta. En la parte de la base, cerca del oído medio, es más angosta que en la parte final o en el ápex, teniendo un ancho 5 veces mayor con 100µm en la base y 500µm en el ápex. Esta variación en la forma, estructura y rigidez de la membrana es lo que le permite a la cóclea analizar las frecuencias. Cuando una vibración se produce en la ventana oval es transmitida al canal vestibular y así mismo esa misma vibración se transmite por la ventana redonda al canal timpánico, lo cual genera una diferencia de presión, desplazando los líquidos linfáticos propagando una onda conocida como onda viajera cuya amplitud depende de la frecuencia del sonido incidente (Soto, Vega, Chávez, & Ortega, 2003). El movimiento de la membrana basilar puede ser pensado como el de una bandera que oscila de acuerdo con lo anteriormente descrito.

En la membrana basilar existen dos tipos de Células Ciliadas, las Internas (CCI) y las Externas (CCE). Estas son responsables de medir la intensidad del estímulo vibratorio y están conectadas con las terminaciones nerviosas. Según (Berenguer Alarcón, i Castellanes, &

---

<sup>15</sup> Este órgano es el equivalente a la retina en el sistema de la visión.

Jordà Puig, 2005) pueden existir hasta 30.000 células entre internas y externas. En donde el artículo de (Soto, Vega, Chávez, & Ortega, 2003) afirma que estas se encuentran ordenadas en una fila de CCI y tres de CCE a lo largo de toda la membrana. Ambos documentos confirman que para cada frecuencia hay una proporción de la membrana que tiene una resonancia de mayor amplitud, la cual activa las CC en una porción única, enviando estímulos bioeléctricos a las neuronas aferentes que hacen sinapsis con estas células. Esta capacidad de discernir frecuencias en proporciones del espacio de la membrana es lo que le permite al cerebro identificar la altura de un sonido: los cambios de frecuencias y sus amplitudes generan cambios en las CC activas en una porción de 3.5 a 4mm. Si la frecuencia se duplica o multiplica por un número arbitrario, el desplazamiento no se multiplica, sino que se mueve en una proporción constante. Lo cual es fantástico, pues en 3.2cm, la membrana organiza un espectro frecuencial logarítmico. Para comprender mejor lo expuesto hasta este punto, ver la figura 5, en donde se muestra el rango de frecuencias audibles y el recorrido de la onda que genera el estribo en la cóclea, así como las zonas que se activan.



*Figura 5 La cóclea y el rango de frecuencias*

*Fuente: tomado de med-el. How hearing works<sup>16</sup>*

De este escueto resumen, cabe decir que hay bastante divergencia entre lo que indica un autor y otro. Mientras que algunos puntos de vista se han solidificado aún hay controversia en cuanto al verdadero sentido de los líquidos linfáticos y en la forma en la cual la membrana basilar logra discernir los estímulos con base en las frecuencias que le transmite el oído medio. Sin embargo, es una idea de lo que sucede en el sistema auditivo y para lo que respecta

<sup>16</sup> Obtenido de: <https://www.medel.com/about-hearing/how-hearing-works>

a la investigación, logra dar cuenta de los parámetros que debe tener un sistema de espacialización del sonido para generar una percepción tridimensional en un ambiente virtual.

- **Etapa 4: Nervio Coclear**

También es conocido como nervio auditivo y es el encargado de conectar las señales bioeléctricas que genera el oído interno con los centros auditivos del cerebro. La energía transmitida desde la cóclea es dispuesta a los dos centros nerviosos de la corteza temporal en ambos hemisferios.

- **Etapa 5: Vías auditivas centrales**

En esta última etapa, el cerebro recibe la información desde las vías centrales y las decodifica. Recibe las características del estímulo y las traduce en una interpretación razonable de lo que significa ese mensaje, relacionándolo con la memoria, la imaginación y la experiencia para generar una apreciación consistente. Es en esta etapa en donde la relación entre el sonido y la percepción sensorial toma sentido, y se proyecta en las reacciones anímicas que cotidianamente hace sentir la música, las voces, el ruido.

### **2.2.1.3 Percepción espacial del sonido y audio binaural**

El ser humano tiene dos oídos para poder diferenciar la procedencia de las distintas fuentes sonoras y para definir su entorno. Esa diferenciación implica un análisis perceptivo de tiempo e intensidad relativos a la fuente sonora, en cuanto a la manera en la que el sonido alcanza el oído. La forma del Pabellón Auditivo, es decir, los múltiples pliegues y formas de la oreja son aspectos fundamentales en el proceso de discriminación de las distintas fuentes sonoras (Soto, Vega, Chávez, & Ortega, 2003). La parte posterior de la oreja funciona como un filtro en las frecuencias bajas, por tanto las circunvoluciones del PA permiten la localización de la fuente sonora en el eje vertical, en donde las frecuencias altas y el espectro en general de los sonidos modifican la percepción espacial. El hecho de tener dos orejas y la cabeza en medio hace que se cree una especie de pantalla acústica que separa las señales que percibe un oído del otro, permitiendo identificar las diferencias de llegada de las ondas acústicas dependiendo de donde se encuentren en el plano horizontal (García Ruiz, 2020).

Para que el sistema auditivo pueda identificar la locación de las fuentes sonoras necesita de ciertas indicaciones. La memoria y la experiencia hacen que ciertos sonidos sean familiares y por tanto fáciles de identificar y ubicar, sin embargo, se requiere de dos pistas sumamente importantes y aún más cuando se requiere identificar sonidos nuevos. La Diferencia de Tiempo Interaural (ITD) y la Diferencia de Intensidad Interaural (ILD)<sup>17</sup> son las principales claves que el sistema auditivo utiliza para localizar las fuentes sonoras en el plano horizontal (acimut). Cuando un sonido se encuentra en frente del oyente (a  $0^\circ$ ), las diferencias son mínimas, al igual que cuando se encuentra atrás (a  $180^\circ$ ), pero estas son máximas si la fuente se encuentra a los lados (a  $90^\circ$ ), ver figura 6. La diferencia de tiempo es aproximadamente igual a la distancia entre los dos oídos<sup>18</sup> dividido por la velocidad del sonido, lo cual en promedio puede estar entre 0.6 y 0.7 ms para cualquier frecuencia.

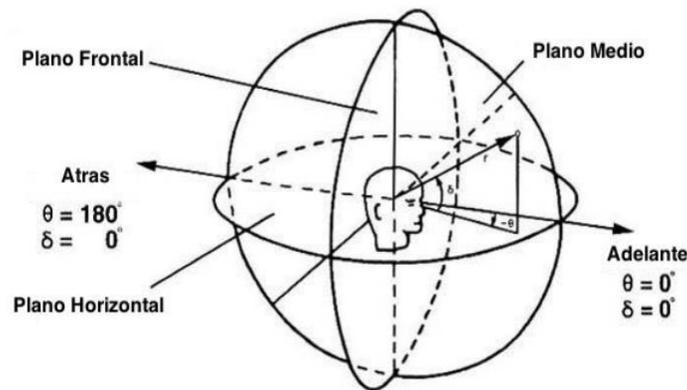


Figura 6 Plano tridimensional para la localización de fuentes sonoras

Fuente: tomado de (Cardona Rojas & Valderrama Pulgarín, 2014)

En el siglo XIX uno de los primeros científicos en realizar estudios en acústica y sonido espacial fue Lord Rayleigh (1842-1910). Quién por medio de un experimento muy conocido, dispuso a un oyente con los ojos cubiertos, en el centro de un círculo rodeado por 5 o 6 asistentes quienes continuamente cambiaban de posición. En el experimento, Rayleigh se dio cuenta que el oyente podía discernir de si una persona estaba a un lado u otro, detrás o delante. Con base en este descubrimiento es que se plantea la teoría dúplex, en donde se establecen

<sup>17</sup> Interaural Time Differences (ITD) – Interaural Level Differences (ILD)

<sup>18</sup> La máxima distancia posible que el sonido debe recorrer para rodear la cabeza es de 21 cm en un ángulo de  $90^\circ$ .

las diferencias de tiempo e intensidad como las pistas clave para localizar un sonido. Estas variables cambian en relación con la frecuencia de la fuente sonora, la sombra acústica generada por el tamaño de la cabeza, las difracciones, forma de los hombros y la oreja junto con otros fenómenos acústicos (Cardona Rojas & Valderrama Pulgarín, 2014).

Las ITD funcionan perfectamente en frecuencias por debajo de los 700Hz en donde las longitudes de onda de la fuente son más grandes que el tamaño de la cabeza. Lo cual hace que la única diferencia entre las señales en los dos oídos sea de fase (ver figura 7). Por encima de 1.6KHz la diferencia de fase es de más de  $180^\circ$  y la longitud de onda menor al tamaño de la cabeza, lo cual hace difícil distinguir cual señal llega primero a un oído u otro. Por tanto, la ITD no es útil para localizar sonidos con frecuencias más altas que 1.5Khz (Zhong, 2015).

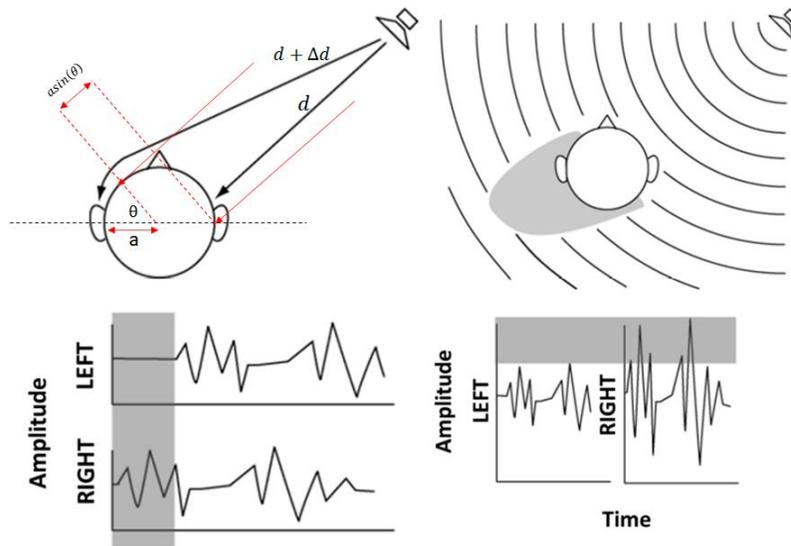


Figura 7 Diferencia de Tiempo Interaural (ITD) y Diferencia de Intensidad Interaural (ILD).

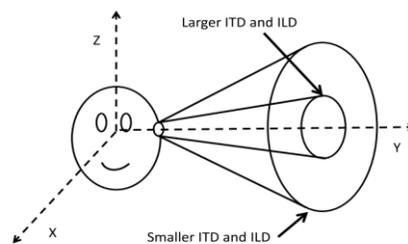
Fuente: tomada y modificada de (Zhong, 2015)

Las ILD lateralizan la fuente con el oído que reciba la mayor intensidad sonora y en principio puede funcionar en todas las frecuencias, pero debido a la sombra acústica de la cabeza<sup>19</sup> no se tiene una gran diferencia de nivel en todas las frecuencias, por lo cual las ILD son más útiles en frecuencias arriba de 1.2KHz (Cengarle, 2012). De lo anterior se encuentra que en el rango de superposición de la ITD con la ILD, la localización no es tan buena

<sup>19</sup> Si la longitud de onda es pequeña la sombra es grande y viceversa. Por tanto, la sombra acústica de la cabeza depende de la longitud de onda de la fuente.

teniendo un mayor aumento de errores. Es en donde el movimiento de la cabeza juega un papel importante, puesto que inconscientemente, el ser humano siempre intentará hacer que las diferencias interaurales sean 0, es decir, haciendo que la fuente sonora este al frente.

Tanto las ITD y las ILD son pistas usadas para interpretar y decodificar la posición de una fuente en el eje horizontal como en el vertical, sin embargo, para el eje vertical el oído utiliza otras pistas para lograr vencer las zonas de ambigüedad en donde las ITD e ILD arrojan los mismos valores. A estas zonas se les conoce como conos de confusión (ver figura 8). La superficie de este cono muestra la imposibilidad de discernir si la fuente se encuentra arriba o abajo del eje horizontal, puesto que los valores de las diferencias interaurales son los mismos. Para ello el sistema auditivo tiene dos opciones, una es el ya mencionado movimiento de la cabeza para que la fuente quede en frente, haciendo que las diferencias sean cero. Pero otra pista que se utiliza para la localización vertical es la fisionomía de la persona<sup>20</sup>, puesto que el contenido espectral de la fuente es modificado por la difracción de los hombros<sup>21</sup>, la forma de la cabeza y el Pabellón Auricular, según la dirección de llegada del sonido. La oreja produce una alteración de la señal de llegada entre los 5 y 10KHz en función de la elevación de la fuente, y esta es la principal manera de localizar un sonido por su elevación, basado en un análisis espectral hecho en tiempo real por el cerebro. Entonces, el oído externo junto con la cabeza y los hombros forman una especie de filtro digital único para cada persona, el cual determina la forma en la que se perciben los sonidos (Cengarle, 2012).



*Figura 8 Cono de confusión*

*Fuente: tomada de (Cengarle, 2012)*

<sup>20</sup> Para la percepción de elevación, el oído utiliza las pistas arrojadas por los rebotes que genera la forma del oído externo. Por tanto, la elevación se modela filtrando los sonidos con respecto a su frecuencia y dirección de llegada. Este filtro se conoce como Comb Filter o filtro tipo peine y es muy usado en filtros digitales.

<sup>21</sup> Efectos entre los 200 y 500Hz.

La percepción de distancia se obtiene mediante 5 mecanismos y el primero es la intensidad. Los sonidos más lejanos llegan con mayor atenuación que los sonidos más cercanos, pero se debe tener una referencia de la potencia de la fuente sonora. Según el tipo de fuente el cerebro estima la potencia de transmisión para así calcular la atenuación. Cuando son sonidos conocidos es más fácil calcular la distancia por medio de la experiencia y la memoria, sin embargo, en sonidos desconocidos es incierto saber la potencia de origen. En situaciones ideales como campo libre o cámaras anecoicas, la relación entre intensidad y distancia se puede calcular por la ley del cuadrado inverso, la cual indica que habrá una reducción de intensidad de 6dB cada que la distancia se duplique (Barrera Giraldo, 2009). El segundo mecanismo es la atenuación en frecuencias altas. La propagación por el aire hace que las frecuencias altas sufran una mayor pérdida de intensidad por fricción con las partículas del aire, así también como con los obstáculos presentes en el medio, los cuales actúan como un filtro. Cuanto más atenuadas se perciban las frecuencias altas de un sonido, es posible que se encuentre más lejos. El tercer mecanismo se trata de paralaje por movimiento. Este es usado con señales de alta frecuencia invisibles o que no se puedan localizar a simple vista. Allí las ITD no funcionan y no se conoce la potencia de origen. Por tanto, se busca un movimiento lateral para que las diferencias de intensidad ILD sean perceptibles. Si la fuente está muy lejos no importa cuánto se desplace la cabeza no se percibirán diferencias de intensidad, pero si el sonido está cerca se logrará percibir intensidades distintas en cada oído. El cuarto mecanismo se trata del exceso de ILD. Esto funciona para fuentes que están muy cercanas a uno de los dos oídos y actúa como una forma de defensa, como en el caso de un mosquito en donde la diferencia es máxima detectando sobre cual oído está la perturbación. Al ser una fuente muy cercana la forma de onda que llega al oído no es plana sino esférica y por tanto aumenta la difracción y la sombra de atenuación. El último mecanismo es la relación entre el sonido directo vs sonido reverberante. Este da cuenta del tamaño del recinto en donde se encuentra la fuente. La señal reverberante es producida por los rebotes del sonido directo en las superficies de la sala. Si el oyente percibe con mayor intensidad el sonido directo que el sonido reverberante, es porque la fuente está cerca. Pero si la intensidad del sonido directo es menor que el de la señal reverberante, la fuente está lejos.

Sobre el audio binaural, es relevante plantear ciertas distinciones. En primer lugar, la diferencia entre audio y sonido radica en su generación y almacenamiento. Por audio se comprende todo archivo de contenido sonoro, cuyo origen ha sido previamente grabado o sintetizado digitalmente y que se encuentra almacenado (o procesado) en una computadora. Mientras que el sonido, se refiere a la forma natural en la cual una perturbación es generada por un objeto físico en vibración el cual existe temporalmente de forma mecánica, como una voz. Por otra parte, el uso de frases alusivas al sonido espacial se refiere al mismo concepto. En adelante se utilizará Audio 3D, pero sonido espacial, sonido 3D, audio binaural 3D, acústica virtual, entre otros, son técnicas que modelan el pabellón auricular como un filtro digital. Filtrando una fuente de audio digital con estos filtros es posible ubicar cualquier sonido en donde sea dentro de un ambiente virtual, usando audífonos. Cualquier modelo o sistema que pretenda sintetizar Audio 3D, sí o sí debe incluir el filtrado del Pabellón Auricular (Begault, 1994).

#### **2.2.1.4 Análisis de sistemas**

Los sistemas Lineales e Invariantes en el Tiempo (LTI) se caracterizan por cumplir dos propiedades fundamentales en el análisis de sistemas, las cuales son la linealidad y la invarianza en el tiempo. Un sistema es lineal cuando cumple la propiedad de superposición, la cual establece que si la señal de entrada es una suma de señales, la salida del sistema será una señal con la respuesta a las distintas señales de entrada. Esto último implica que el sistema cumpla la propiedad de aditividad. Por otra parte, un sistema lineal también debe cumplir la propiedad de escalabilidad u homogeneidad, la cual implica que si la señal de entrada es proporcional a una constante compleja cualquiera, la respuesta del sistema también será proporcional a esa constante. Finalmente, un sistema es invariante en el tiempo cuando se pueden obtener las mismas respuestas tiempo después de haber hecho una medición inicial, siempre y cuando se mantengan iguales los parámetros que intervienen en esa respuesta. Por tanto, dicha respuesta debe permanecer fija en el tiempo para obtener los mismos resultados. Si los parámetros cambian, es de esperar que la respuesta del sistema sea distinta, sin embargo, en sentido estricto la invarianza en el tiempo hace que una señal de entrada con corrimiento en el tiempo implique una señal de salida corrida en ese mismo tiempo.

Un concepto fundamental para construir un entendimiento básico sobre las Funciones de Transferencia Relacionadas con la Cabeza (HRTF), es el impulso unitario, el cual se refiere a una función a trozos cuyo único valor distinto a cero, se da en un tiempo igual a cero. En donde su valor en el tiempo continuo puede ser infinito, y en tiempo discreto es igual a 1. De allí que la característica más importante del impulso unitario es que puede reconstruir cualquier señal por medio de la combinación lineal de impulsos retardados. Esto último unido a las propiedades de superposición e invarianza en el tiempo, permiten caracterizar el comportamiento de cualquier sistema LTI en términos de su respuesta al impulso (Oppenheim & Willsky, 1998).

La convolución es una función de gran importancia cuando se trabaja con procesamiento digital de señales y es la clave para comprender el funcionamiento del audio 3D. La convolución transforma una señal de entrada a un sistema (con una función de transferencia y una respuesta al impulso conocida), en la forma en la cual ese sistema respondería a esa señal de entrada, obteniendo de manera analítica el comportamiento de esa señal en el sistema. Por ejemplo, un eco es la convolución de un sonido directo con la función de transferencia de una sala con superficies reflectantes. Así pues, para tener la sensación de que un sonido proviene desde una ubicación particular en un mundo virtual, se requiere convolucionar la fuente de sonido con las HRTF de cada oído. Usando audífonos el resultado será una reproducción bastante aproximada de la posición espacial de esa fuente sonora (ver figura 9 izquierda).

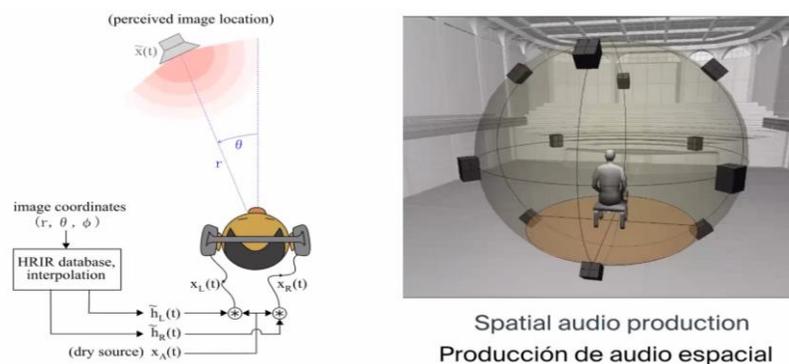


Figura 9 A la izquierda el sistema HRIR-oyente. A la derecha una representación de audio espacial con altavoces.

Fuente: tomado de la ponencia de Marcos Simón en el 1er simposio sobre parlantes y sus aplicaciones.<sup>22</sup>

<sup>22</sup> Información del evento en: <https://www.acusticauach.cl/?p=14022>

La transformada de Fourier es una función extremadamente útil en el procesamiento de audio moderno. A tal punto que ningún analizador de espectro digital o ecualizador de audio digital se hubiese podido crear, sin la implementación de las transformadas de Fourier. El concepto principal aquí, se trata de la descomposición de cualquier tipo de señal en una sumatoria de señales sinusoidales de magnitudes y frecuencias distintas. En el campo discreto, esta descomposición se puede realizar mediante impulsos, lo cual facilita el análisis de los sistemas y permite aplicar a ellos las operaciones de convolución (Barrera Giraldo, 2009).

De lo anterior, es posible pensar el sistema auditivo como un sistema LTI si se toman posiciones fijas de la fuente sonora y los oídos. Por tanto, para conocer su comportamiento basta con conocer la respuesta al impulso que genera la fuente localizada dentro de un espacio tridimensional cualquiera. Luego obtener la transformada de Fourier de esa respuesta al impulso y convolucionar esa función de transferencia resultante para obtener un sonido virtualmente espacializado.

#### **2.2.1.5 Funciones de Transferencia Relacionadas con la Cabeza (HRTF)**

El audio 3D busca trasladar las sensaciones acústicas de un entorno real, natural y original a uno virtual, y para lograrlo existen varias técnicas. Como se indicó arriba, se puede medir la presión acústica de un campo sonoro en la posición de un oyente, y luego se puede reproducir la misma señal en los oídos del oyente para que la información original sea reconstruida. Esta es la forma tradicional de medir la respuesta de cada persona ante una fuente sonora y se utiliza en laboratorios con cámaras anecoicas. Allí se toman los datos de las respuestas al impulso generando una esfera alrededor del punto de escucha<sup>23</sup> (ver figura 10). Debido a la estructura anatómica única de cada humano se tiene una HRTF individual para cada oído. Por ello, se pueden obtener HRTFs únicas pero el proceso es costoso y no es tan viable para Realidad Virtual en donde con HRTFs genéricas es posible trabajar con las ambigüedades. Por este motivo, se ha demostrado que además de las diferencias en las medidas de HRTF de los individuos, hay algunos patrones básicos que funcionan para todos los seres humanos. Estos son conocidos como bandas direccionales. Y se refieren básicamente a incrementos y atenuaciones en frecuencias relacionadas con las posiciones de

---

<sup>23</sup> Para obtener la respuesta del oyente, se utilizan arreglos de altavoces y micrófonos binaurales

los sonidos. Las regiones alrededor de 300 a 600Hz y de 3 a 6 KHz parecen estar relacionadas con las posiciones frontales del PA, las de 8 KHz parecen corresponder a la posición sobre la cabeza y las áreas de 1,2 KHz y 12 KHz parecen estar relacionadas con la percepción posterior. Esta es la razón por la que la reproducción binaural a través de auriculares utilizando HRTFs genéricas funciona de manera exitosa para la mayoría de personas (Systematic Sound, 2020). Sin embargo, gracias a las mediciones en cámaras anecoicas se han creado bases de datos<sup>24</sup> que actualmente son utilizadas en los modelos matemáticos que logran engañar al sistema auditivo. Con esas bases de datos se pueden sintetizar las HRTFs por convolución. De esta manera se pueden reproducir escenarios realistas para Realidad Virtual, los cuales también permiten el seguimiento de la cabeza conocido como Head Tracking, el cual adapta la localización del sonido con respecto al movimiento de la cabeza (Cengarle, 2012).

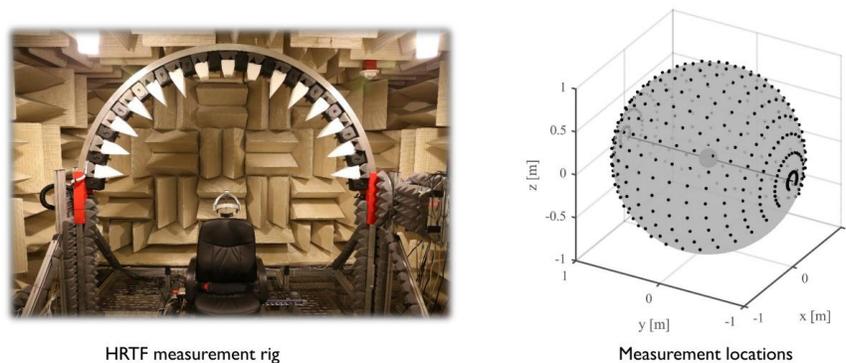


Figura 10 Cámara anecoica para medición de HRTF y esfera obtenida con las respuestas al impulso

Fuente: [https://signalprocessingsociety.org/sites/default/files/Audio\\_challenges\\_in\\_virtual\\_and\\_augmented\\_reality\\_devices\\_-\\_IWAENC\\_-\\_fin....pdf](https://signalprocessingsociety.org/sites/default/files/Audio_challenges_in_virtual_and_augmented_reality_devices_-_IWAENC_-_fin....pdf)

Las HRTF modelan todos los efectos acústicos que se producen en una fuente sonora desde su origen hasta la llegada al interior del oído para simular la forma en que se percibe el sonido. Por tanto, las HRTF codifican las diferencias interaurales y las estructuras anatómicas para la locación de las fuentes sonoras en un ambiente virtual, y para ello utilizan los ángulos de azimut y elevación (ver figura 11) con el que incide el sonido en el oído (Cardona Rojas & Valderrama Pulgarín, 2014). Por otra parte, el audio 3D utiliza la

<sup>24</sup> Como las de IRCAM <http://recherche.ircam.fr/equipements/salles/listen/>. Las de Media Lab - MIT <http://sound.media.mit.edu/resources/KEMAR.html>. Las de CIPIC <http://interface.cipic.ucdavis.edu/sound/hrtf.html>.

reproducción binaural, la cual requiere de altavoces<sup>25</sup> o audífonos para entregar dos señales correspondientes a cada oído<sup>26</sup>. En ese sentido, deben existir dos HRTFs tanto para el canal derecho como el izquierdo. Además, la principal diferencia entre el audio 3D y el estéreo está en la externalización de la fuente sonora, esto es, la profundidad en cuanto a la atenuación de distancias entre el oyente y el origen del sonido (Brinkman, Hoekstra, & Egmond, 2015). Por tanto, en experiencias inmersivas es obligatorio que el audio 3D se perciba fuera de la cabeza de quien lo recibe, y esto implica atenuaciones en función de la distancia para cada HRTF.

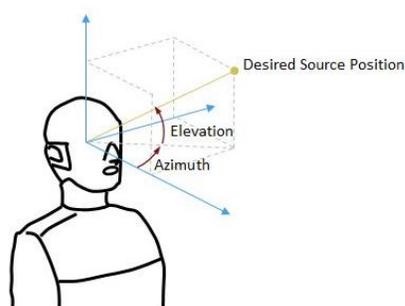


Figura 11 Parámetros básicos de la HRFT.

Fuente: [www.wikiversus.com/audio/hrtf-sonido-posicional/](http://www.wikiversus.com/audio/hrtf-sonido-posicional/)

Las diferencias interaurales ITD - ILD, se pueden modelar como un arreglo de dos micrófonos separados por una distancia fija, por ello existen maniqués con solo la forma de la cabeza y hombros con dos micrófonos dentro de las orejas. Normalmente suelen utilizar micrófonos omnidireccionales y se usan para hacer las pruebas en las ya mencionadas cámaras anecoicas, grabación binaural o para captar paisajes sonoros como el caso del “dummy head: Neumann KU-100”<sup>27</sup>. Para más información sobre grabación y micrófonos binaurales, ver el documento de (Wittek, 2016)<sup>28</sup>.

<sup>25</sup> Audioscenic es una empresa que fabricó un arreglo de altavoces en una barra de sonido, la cual por visión artificial y filtros de cancelación cross-talk permiten recibir audio completamente 3D. <https://audioscenic.com/>

<sup>26</sup> Para una mayor inmersión se recomienda siempre usar audífonos.

<sup>27</sup> <https://en-de.neumann.com/ku-100>.

<sup>28</sup> O visitar [www.schoeps.de](http://www.schoeps.de), incluso en <https://hookeaudio.com> se encuentra tecnología de punta enfocada en micrófonos de grabación binaural.

Finalmente, en la figura 12 se resume todo el proceso que debe ser llevado a cabo para obtener una fuente sonora completamente espacializada mediante HRTFs. Cabe mencionar que actualmente las investigaciones y productos comerciales se han enfocado en conseguir HRTFs personalizadas para cada persona. Estas se consiguen mediante visión artificial en donde se computan las medidas del PA para obtener filtros únicos<sup>29</sup>.

Las aplicaciones del audio 3D son variadas y cada día van a existir mejores métodos para obtener HRTFs personalizadas. Sin embargo, el campo en el cual se han podido aplicar los avances hasta el momento es en la Realidad Virtual. Desde los inicios de esta tecnología se tenía previsto que existieran gafas o cascos con una pantalla en el interior, junto con un par de audífonos a los lados. Los cuales hicieran posible que el sonido fuera procesado en tres dimensiones, para que el oyente tuviera la percepción de que los sonidos vinieran desde ciertas direcciones (Lanier, A Vintage Virtual Reality Interview, 1988).

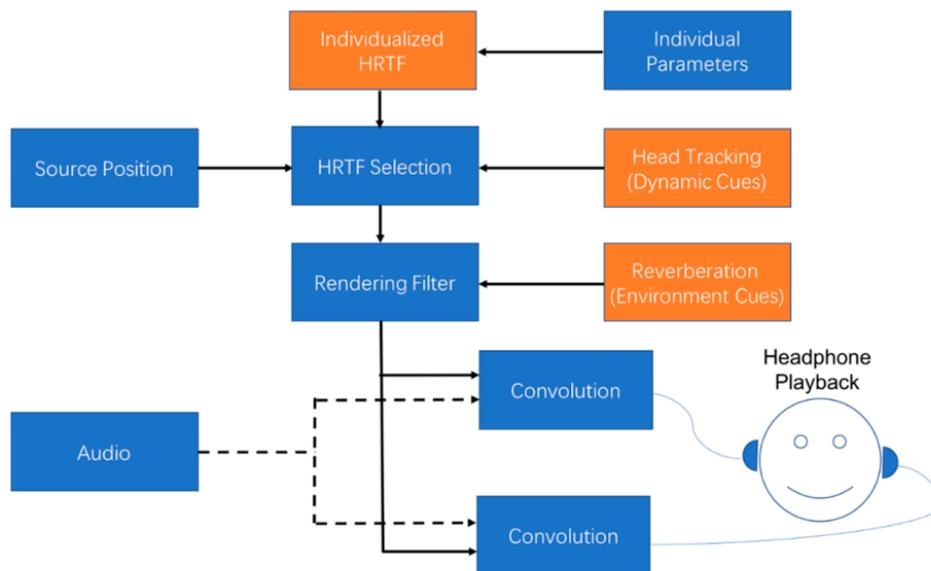


Figura 12 Flujo de la señal de un sistema de reproducción binaural.

Fuente: tomado de (Zhang, Samarasinghe, Chen, & Abhayapala, 2017)

<sup>29</sup> La empresa Steinberg ha lanzado un nuevo VST que permite tomar una foto a la oreja y obtener un perfil con una HRTF personalizada <https://new.steinberg.net/immerse/>

### 2.2.2 Realidad Extendida

Es complejo encontrar una definición estable para una tecnología que se transforma<sup>30</sup> con una fuerza evolutiva tan voraz. Cuyos términos y dispositivos permanecen en constante ampliación mediante nuevas herramientas y métodos de trabajo. Cuyas líneas divisorias cada vez se desdibujan más. A este respecto es pertinente la siguiente inquietud, ¿Cómo definir conceptualmente algo que no ha sido delimitado de manera explícita y que sobre su base más primitiva de existencia reside la idea de que por sí misma no tenga límites? Sin embargo, es relevante atreverse a buscar tierra firme sobre la cual se pueda reflexionar y encontrar ideas comunes, por tanto, antes de empezar a identificar lo que es la Realidad Extendida (conocida como XR), es justo mencionar una mínima información de contexto.

La Realidad Virtual fue lo que primero se definió y fue gracias al esfuerzo de varias personas en distintos tiempos. La historia como tal, se puede remontar a las primeras imágenes panorámicas en 360 grados<sup>31</sup>, sin embargo, los nombres de Ivan Sutherland y Jaron Lanier retumban porque se les atribuye ser los creadores del término Realidad Virtual. Sutherland<sup>32</sup> fue el primero en crear un display para ser montado en la cabeza y Lanier fue uno de los fundadores de la primer empresa<sup>33</sup> que comercializó dispositivos para Realidad Virtual e inventó un lenguaje de programación y toda una taxonomía de lo que sería esta tecnología (Lanier, A Vintage Virtual Reality Interview, 1988). Junto con Tom Zimmerman<sup>34</sup> vendían todo un sistema que involucraba guantes de datos, cascos con pantallas montadas en la cabeza y computadoras especiales las cuales gestionaban todo el procesamiento de gráficos, audio y seguimiento (ver figura 13). De esta empresa surgió el primer sistema de audio 3D para RV, llamado Audio Sphere.

---

<sup>30</sup> “Y que nos transforma” (aquí me disculpo por faltar a la escritura del texto en tercera persona)

<sup>31</sup> <https://www.vrs.org.uk/virtual-reality/beginning.html> Una obra sobre la historia de la Realidad Virtual a un nivel profundo podría ser un digno tema de investigación para el proyecto de investigación ALICE. En la página de la Virtual Reality Society se encuentra bastante información que puede ser un comienzo para un trabajo serio, al igual que en <https://hookeaudio.com/blog/virtual-reality/history-of-virtual-reality/>, así también como en la literatura de cabecera.

<sup>32</sup> Inventor del primer Headset o eyeglasses, también conocidos como Head-Mounted Displays (HMD).

<sup>33</sup> Aproximadamente en 1984 se crea VPL Research - <https://www.vrs.org.uk/virtual-reality-profiles/vpl-research.html>

<sup>34</sup> Inventor de los guantes de datos usados para la percepción háptica.

Por otra parte, Lanier es un científico informático, compositor, artista visual y autor de varias obras literarias en las cuales ha discutido sobre temas profundos en cuanto a la filosofía de la tecnología, su futuro y actualidad <sup>35</sup>.

La Realidad extendida engloba lo real y lo virtual, en donde existe una serie de conjuntos o clasificaciones que ayudan a comprender el campo de acción y los límites de una serie de tecnologías que se aplican para percibir lo real y lo virtual de maneras diferentes. La Realidad Extendida es un término poco generalizado que ha surgido en los últimos años para tratar de sugerir que existe un extremo en donde todo estímulo sensorial es real, otro extremo donde todo estímulo sensorial es creado digitalmente por medio de gráficos por computadora y una instancia en el medio que puede tener una fusión de los dos extremos. La Realidad Extendida es un término que busca abarcar todo el Continuo de la Realidad-Virtualidad definido por (Milgram, Takemura, Utsumi, & Kishino, 1994) incluyendo la compleja línea que divide las categorías descritas y a su vez busca englobar las nuevas tecnologías que surjan en adelante. En ocasiones se suele nombrar a todas estas tecnologías simplemente como Realidad Virtual, pero hacerlo con lleva a que las definiciones sean limitadas y por ende sea imperante encontrar un marco conceptual más amplio.



The first virtual reality system, according to the original definition, in which multiple people cohabited a virtual world at the same time. This was VPL's RB2, or "Reality Built for Two." In the screens behind each person, you can see how they see each other as avatars. This photo is from a trade show in the late 1980s.

*Figura 13 Primeros dispositivos de Realidad Virtual de la empresa VPL*

*Fuente: tomado de (Lanier, Dawn of the New Everything: A Journey Through Virtual Reality, 2017)*

---

<sup>35</sup> Se recomiendan textos como, Ten Arguments For Deleting Your Social Media Accounts Right Now. You Are Not a Gadget. Who Owns the Future?. Dawn of the New Everything: Encounters with Reality and Virtual Reality, entre muchos más, siendo estos los que mejor describen su pensamiento y activismo filosófico en la actualidad.

La línea del Continuo de la Realidad-Virtualidad en Milgram et al. (1994) sostiene que los extremos no son la antítesis de uno ni del otro, sino que son simplemente extremos opuestos en una línea con dispositivos tecnológicos que modelan la percepción del usuario, según el grado en que los sentidos estén inmersos en un ambiente controlado por graficas de computadora o por las leyes de la física que gobiernan el mundo real y natural. Unificar el concepto del Continuo de la Realidad-Virtualidad de estos autores con los nuevos planteamientos de la Realidad Extendida, es un hito para la conceptualización de este tipo de tecnología desde una postura completamente científica y técnica.

En el centro de la Realidad Extendida, se tienen dos categorías llamadas Realidad Mixta, las cuales mezclan la realidad y lo virtual de distintas maneras. Una de estas categorías se conoce como Realidad Aumentada (RA) y busca superponer objetos gráficos creados por computadora sobre estructuras reales. Para ello se usan cascos con pantallas transparentes que puedan mostrar o renderizar los objetos virtuales<sup>36</sup>, o también se usan pantallas que por medio de una cámara superponen los gráficos y son conocidas como Window-on-the-World (WoW) (Milgram, Takemura, Utsumi, & Kishino, 1994). Esta última es muy usada en aplicaciones para dispositivos móviles.

Por otro lado, se encuentra la categoría de la Virtualidad Aumentada en donde el mundo alrededor del usuario es principalmente virtual, pero es aumentado por objetos reales. Un ejemplo de este tipo de categoría se observa en la tecnología de Hand Tracking<sup>37</sup> de las gafas de Realidad Virtual Oculus Quest, en donde las manos hacen parte de la realidad y tienen una representación en el mundo virtual puesto que pueden manipular objetos virtuales sin la necesidad de controles. En la investigación elaborada por (Quevedo Rodríguez & Malagón Barrera, 2019) se evidencia un ejemplo de Virtualidad Aumentada, en donde algunos videos 360 y un extintor reales, son incluidos dentro de una aplicación para Realidad Virtual, la cual sirve de guía para explicar el manejo de extintores.

De lo anterior, es posible dilucidar el nivel de complejidad para definir una categoría de la otra, puesto que en ocasiones no es obvia la identificación inmediata sobre si el entorno principal es real o virtual, y los cambios acelerados en estas tecnologías apuntan a que las

---

<sup>36</sup> Como las gafas HoloLens. <https://www.microsoft.com/en-us/hololens>

<sup>37</sup> Ver la información del fabricante a este respecto. <https://support.oculus.com/2720524538265875/>

distinciones sean cada vez más difíciles de limitar, teniendo como último recurso, los conceptos que puedan ser encontrados en la Realidad Mixta y la Realidad Extendida.

### 2.2.2.1 Realidad Virtual

Con lo dicho hasta ahora, la Realidad Virtual (RV) es el extremo opuesto a la realidad material, visto desde los conceptos planteados por Milgram et al. (1994) y por (Rubio Tamayo, 2019). Allí, en donde el entorno es netamente virtual, todo puede ser posible (Lanier, Dawn of the New Everything: A Journey Through Virtual Reality, 2017)<sup>38</sup>. En ese extremo del Continuo, las leyes de la física pueden ser distintas o no existir. La representación en forma de avatar del usuario puede ser la misma que en la realidad<sup>39</sup> o puede ser totalmente diferente. Lanier en los 80's aseguraba que en la Realidad Virtual el usuario podía ser un gato o un instrumento musical, estar sentado en una silla de un museo lejano o en los anillos de saturno (Lanier, A Vintage Virtual Reality Interview, 1988). En esa misma entrevista Lanier planteaba conceptos que son vigentes en la actualidad y es contundente al momento de afirmar que la Realidad Virtual no se puede pensar como una sucesión evolutiva en los computadores y por tanto, su futuro no se puede ligar a las limitaciones de una herramienta de mesa como es el Laptop o el Desktop, confrontándolo con los dispositivos de Macintosh. Sino que se debe pensar como un medio, una tecnología que comunica a las personas en la cual la inmersión es sinónimo de expresión, de presencia, de externalización y conservación de las experiencias y la memoria.

La Realidad Virtual integra la narrativa y la experiencia del usuario, por medio de una interfaz que supone el intercambio de estímulos entre los sentidos y las acciones que modifican el entorno virtual. A esto se le conoce como inputs, mientras que la forma en la cual estos estímulos influyen en el comportamiento del usuario por los medios gráficos computacionales se le conoce como outputs (Rubio Tamayo, 2019). De lo anterior, existen unos elementos básicos que todo sistema de Realidad Virtual debe contemplar. La simulación de entornos tridimensionales con interacción implícita y la inmersión sensorial son objetivos imperativos a conseguir. La simulación interactiva es lo que diferencia a la Realidad Virtual

---

<sup>38</sup> Al menos desde sus postulados filosóficos.

<sup>39</sup> Existen avatares basados en fotografías, así también como otros hechos con base a una representación en anime como los Vtubers. <https://www.realovirtual.com/articulos/5612/evolucion-los-avatares-realidad-virtual-facebook-oculus>

de una animación, puesto que el usuario puede percibir el entorno establecido, experimentarlo y modificarlo si este permite que los objetos puedan ser transformados, ya sea en su volumen o posición tridimensional. La interacción implícita por su parte, apunta a que las interacciones dentro de la Realidad Virtual se den de manera natural, con movimientos corporales idénticos a los que en la realidad un ser humano realiza para mover objetos o desplazarse. La diferencia con otros tipos de interacción esta en las acciones que el usuario tiene que llevar a cabo para realizar algún cambio dentro del entorno. Por ejemplo, memorizar una secuencia de botones para desplazarse, girar la cabeza o accionar alguna función. En Realidad Virtual por ejemplo, para girar la cámara y reconocer el entorno basta con mover el cuerpo o la cabeza hacia la dirección que se quiera visualizar, sin necesidad de memorizar una secuencia de acciones. La inmersión sensorial es la desconexión de los sentidos principales<sup>40</sup> del mundo real hacia una conexión con el entorno virtual. El principal factor de inmersión se da por medio de la visión, en donde la Realidad Virtual por medio del HMD, implementa algo conocido como visión estereoscópica, por medio del cual es posible reproducir la percepción de profundidad, de distancia de los objetos con respecto al usuario, además de la sensación de estar en un espacio tridimensional. Por ello las gafas de RV utilizan dos pantallas, una para cada ojo. También, se utilizan los métodos descritos con anterioridad sobre sonido espacial para hacer que las fuentes sonoras tengan una posición específica dentro del entorno simulado (Pérez, 2015).

La Realidad Virtual inicialmente fue descrita para que funcione mediante un computador el cual se encarga de procesar en tiempo real los gráficos, el audio espacial, las interfaces de usuario – máquina y toda la programación que se requiere para que todo lo anterior funcione en sincronía dentro de un ambiente virtual tridimensional. Requiere de un Head-Mounted Display (HMD) o también conocido como gafas de RV. Allí se pueden encontrar una o dos pantallas, las cuales tienen que tener una frecuencia de muestreo no menor a 90Hz para poder garantizar una buena y sana experiencia inmersiva<sup>41</sup>. También se tienen guantes de

---

<sup>40</sup> Como la visión y la audición.

<sup>41</sup> Existe un problema con la Realidad Virtual llamado Motion Sickness, el cual genera náuseas en las personas que usan las gafas por un tiempo. Esto puede ser generado por una baja tasa de muestreo en los lentes haciendo que el cerebro se desconecte de la inmersión y genere estímulos de alerta que indican al cerebro que algo anda mal. Otro aspecto que causa náuseas y mareo es el desplazamiento por medio de los entornos, puesto que en cierta forma el usuario se desplaza en lo virtual pero su cuerpo físicamente está de pie en un solo punto. Esto es interpretado por el cerebro como una anomalía y su forma de atacar el problema es generando el malestar

retroalimentación háptica o controles que permiten la interacción con el entorno e interfaces gráficas. En algunos dispositivos existen sensores externos que por medio de infrarojos detectan la posición del HMD y de los controles para hacer un seguimiento sincronizado con la aplicación en el computador. Sin embargo, las nuevas gafas de la empresa Oculus, adquirida por Facebook, ya no requiere de sensores externos puesto que el HMD incorpora cuatro cámaras que sirven para hacer el monitoreo y también incluyen el seguimiento de manos por lo cual no es del todo necesario el uso de los controles (ver figura 14).



*Figura 14 Gafas Oculus Quest 2*

*Fuente: <https://www.oculus.com/quest-2/>*

### **2.2.2.1.1 Oculus Quest**

Cabe mencionar desde ya, que este proyecto fue desarrollado para dispositivos Oculus Quest, y aunque en un comienzo se haya planteado para las Oculus Rift<sup>42</sup>, estas gafas tienen varias peculiaridades que hicieron viable la realización de la aplicación en esta plataforma. La primera mejora con relación a los anteriores dispositivos de Realidad Virtual es que estas gafas no necesitan conexión directa a un computador. Su sistema operativo está basado en Android por lo cual es posible que las aplicaciones sean almacenadas en las mismas gafas como si se tratase de un celular smartphone, por ello se consiguen versiones de 64 y 128 GB de almacenamiento. Por otra parte, como ya se mencionó, estas gafas no requieren de sensores externos puesto que cuentan con 4 cámaras ubicadas en la parte frontal, estas funcionan para hacer el seguimiento de controles y de manos en tiempo real, así también

---

descrito. Por ello, existen varios tipos de desplazamiento para evitar el mareo por movimiento. <https://www.vrs.org.uk/virtual-reality/how-does-it-affect-us.html>

<sup>42</sup> Son el modelo anterior a las Quest. <https://www.oculus.com/rift/>

como para poder ver el exterior mediante la función de Passthrough<sup>43</sup>. Otra mejora de las Quest se refiere a la posibilidad de conexión con un computador por medio de un cable USB<sup>44</sup> 3.0 (o de mayor capacidad de transmisión de datos). Esto permite tener acceso al Home de Oculus Rift y a su vez permite hacer pruebas de desarrollo en tiempo real sin cargar las aplicaciones en forma de APK y usar la memoria de las gafas, por tanto, los gráficos y procesamiento en general es realizado por la computadora.

En la figura 14 se presentó una versión reciente de las Oculus Quest la cual salió a la venta en la segunda mitad del año 2020, mientras que las Quest “normales o Quest 1” salieron el 21 de mayo de 2019. Las diferencias son notorias en cuanto a gráficos y tamaño pero el mayor cambio se debe a su precio de lanzamiento; las Quest 2 salieron al mercado por 299 dólares mientras que las Quest tuvieron un precio de 399<sup>45</sup>.



*Figura 15 Gafas Oculus Quest en Home Studio.*

*Fuente: elaboración propia.*

## **2.2.3 Desarrollo de software**

### **2.2.3.1 Motor de Videojuego**

El motor de juego representa el núcleo de un videojuego y por tanto determina la manera en la cual funcionan los distintos módulos que hacen parte de un desarrollo. Un motor de videojuegos está pensado para hacer más rápido el flujo de trabajo y para la reutilización de material ya existente. Es una herramienta que evita trabajar desde cero bajo la complejidad

<sup>43</sup> La cámara externa permite salir temporalmente de las aplicaciones de Realidad Virtual para observar el exterior. <https://support.oculus.com/609948326077538/>

<sup>44</sup> A esta función se le conoce como Oculus Link. Y se vende por separado <https://support.oculus.com/444652019523821>

<sup>45</sup> Ambos precios para las versiones de 64GB.

de procesos de bajo nivel dotando a los desarrolladores de un entorno orientado en su arquitectura, hacia la reutilización de software y la personalización del motor según las necesidades específicas del juego o aplicación en desarrollo (Vallejo Fernández, y otros, 2014). Un motor de video juego es un software que facilita el cálculo de modelos gráficos y comportamientos físicos en particular. Permite de manera fácil el acceso a ficheros y la incorporación de distintos tipos de formatos de audio, imagen, código y modelos 2D-3D (Lidon, 2019).

La función principal de un motor de videojuegos es servir de base para la creación de aplicaciones que requieran simular entornos por medio de gráficos renderizados, por tanto, cualquier motor debe tener un módulo de render que permita visualizar gráficos creados por computadora en tiempo real. Así también como módulos de sonido, de físicas, de interfaces para el control de usuario, animaciones, detección de colisiones y de inteligencia artificial entre otros (ver figura 16). La mayoría de motores permiten la importación de archivos o herramientas de terceros ya que es más económico y se obtienen mejores resultados. Todo esto en última instancia favorece al desarrollador puesto que puede realizar un trabajo eficiente, de alto nivel en tiempos relativamente cortos.

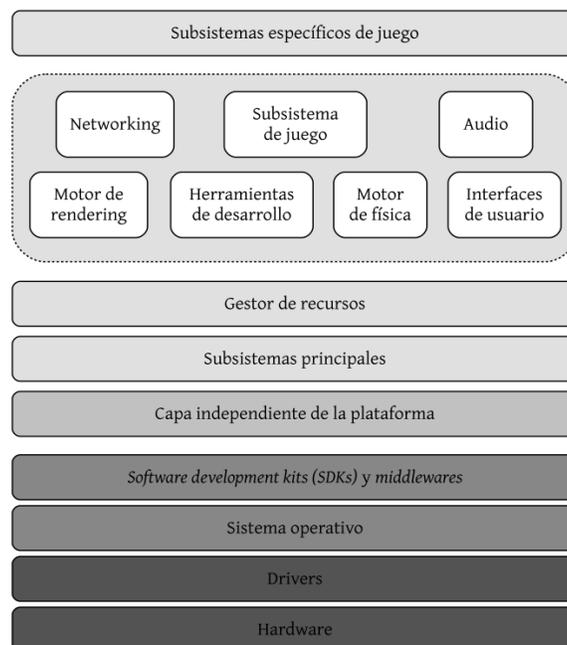


Figura 16 Arquitectura general de un motor de juegos.

Fuente: Obtenido de (Vallejo Fernández, y otros, 2014)

Por último, la mayoría de motores permiten el desarrollo de software para casi cualquier tipo de plataforma. Lo que da mayor libertad al programador para que pueda fluir en el prototipo sin pensar en la plataforma final, puesto que el motor permite exportar el juego o aplicación para la plataforma que se desee, desde PC, MacOS, Android, Xbox, etc.

Los principales motores de videojuegos son Unity, Unreal Engine, Frostbite Engine, Source, CryEngine. Este último es de código abierto por lo cual es de uso libre, mientras que los dos primeros tienen versiones gratuitas para uso académico y personal, o licenciado si se planea un uso comercial que exceda un monto de ganancias específico. En el caso de Unity si una empresa registra ganancias de más de 100.000 dólares anuales, deja de ser gratuito y se debe pagar mensualmente \$200.00 USD<sup>46</sup>.

Cada motor de videojuegos tiene funciones y características que lo hacen especial para cierto tipo de aplicaciones. Por ejemplo Unreal provee gráficos hiperrealistas, mientras que Unity se ha convertido en el principal motor para Realidad Virtual. Por ello, el presente desarrollo fue realizado en ese motor.

### **2.2.3.1.1 Unity 3D**

Unity 3D es un motor que permite un trabajo ágil y eficaz mediante la incorporación de herramientas de terceros y propias. Es un software que a lo largo de los últimos años se ha posicionado como un sólido entorno para el desarrollo de videojuegos multiplataforma de alta calidad<sup>47</sup>. Fue creado por Unity Technologies en el 2005 y desde entonces existen versiones por año con distintas compilaciones que resuelven errores o bugs. Una de las grandes ventajas de trabajar con Unity es la comunidad que hay detrás, puesto que en la web se consigue información de casi todo lo relacionado con sus módulos y resolución de problemas. Cuenta con documentación detallada del funcionamiento de absolutamente todo el software, tanto en la interfaz de edición como en el API<sup>48</sup>. Cuenta con dos sitios en su página web de uso constante como el Asset Store<sup>49</sup>, en donde se pueden comprar Assets o

---

<sup>46</sup> <https://store.unity.com/compare-plans>

<sup>47</sup> Algunos videojuegos hechos en Unity son: Asasin's Creed Identity, Resident Evil: Umbrella Corps, Lara Croft Go, HearthStone, Pokemon Go, Rust, Super Mario Run, entre otros. Y para Realidad Virtual, Job Simulator, The Lab y Rick and Morty.

<sup>48</sup> Application Programming Interface.

<sup>49</sup> <https://assetstore.unity.com/> Los Assets son paquetes con recursos que se pueden implementar de manera personal. Cuando se compran, se puede hacer uso comercial de la licencia. De igual manera si estos son gratuitos en su mayoría tienen una licencia de uso libre. La mayoría vienen documentados o se puede

descargarlos de forma gratuita. También la parte de Unity Learn<sup>50</sup> esta provista de tutoriales y recursos para aprender conceptos y trucos de primera fuente. La mayoría se encuentran en inglés pero cuentan con traducciones al español. Es imperativo que se revise la documentación de la página oficial ya que provee todo lo necesario para resolver problemas o simplemente ingresar al mundo del desarrollo de videojuegos.

Unity permite importar formatos de modelos 2D y 3D como 3ds Max, Blender, Cinema 4D, FBX, OBJ, entre otros. Permite importar audios en formato WAV y en MP3, imágenes en JPG, BMP y PNG, así también como ficheros .TXT. Cuenta con sus propias extensiones de archivos para las Escenas<sup>51</sup>, Animaciones y Texturas. Las Escenas son los archivos principales en Unity y es en donde se almacenan los objetos que hacen parte de un entorno (Lidon, 2019). Se pueden tener varias Escenas en un mismo proyecto de Unity y se pueden cargar Escenas aleatoriamente desde cualquier otra Escena. El concepto de Escena es fundamental para entender los niveles de un juego o los cambios de lugar que se pueden crear.

Los lenguajes de programación fundamentalmente ayudan a modificar el comportamiento de los objetos que se crean dentro de las Escenas. A esos objetos se les conoce como GameObjects y son los elementos principales en Unity pues representan cualquier cosa dentro de un entorno. Estos GameObjects también sirven como contenedores de Componentes, los cuales modifican el comportamiento o las funciones de esos objetos en sí mismos, o el conjunto de objetos que conforman una Escena, así también como el mismo comportamiento del videojuego. Las Escenas, los GameObjects y los Componentes son los bloques fundamentales de trabajo en Unity, de estos se derivan la mayoría de procesos y métodos de desarrollo.

Otro concepto clave en el manejo de Unity es el de Prefab, el cual se refiere a un fichero que almacena uno o varios GameObjects junto con sus componentes y parámetros para ser reutilizados en la misma Escena o en otras, incluso en otros proyectos. Los Prefabs ayudan bastante a economizar tiempo puesto que se pueden repetir objetos sin la necesidad de volver

---

contactar directamente al creador. En el store hay recursos de toda clase, desde modelos 3D, texturas, sonidos, videos, códigos de programación, interfaces de usuario, animaciones, proyectos, etc.

<sup>50</sup> La mejor manera de comenzar con Unity es aquí: <https://learn.unity.com/>

<sup>51</sup> Por ejemplo “.unity”.

a crearlos, modificarlos o parametrizarlos. En la figura 17 se encuentra un esquema que resume los conceptos explicados.

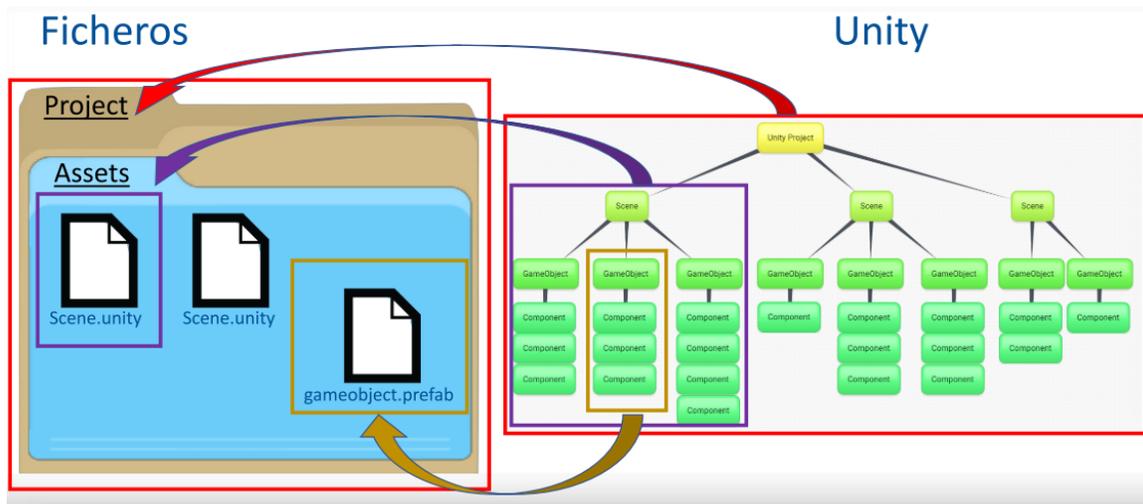


Figura 17 Términos Clave en Unity 3D

Fuente: obtenido de <http://diegomrivera.com>

### 2.2.3.2 Visual Studio y C#

Unity funciona en conjunto con otros programas que se encargan de gestionar las librerías y los códigos de programación. A este tipo de software se le conoce como IDE (Integrated Development Environment) y son aplicaciones con servicios integrales que le facilitan el trabajo al desarrollador. Normalmente, están compuestos de un editor de código, un compilador o constructor automático, un intérprete y un depurador, los cuales sirven para verificar los códigos y ponerlos en funcionamiento. La mayoría de IDE tienen funciones de IntelliCode, las cuales ayudan a autocompletar líneas de código, lo cual agiliza la escritura y evita cometer errores.

El IDE utilizado en conjunto con Unity es Visual Studio<sup>52</sup>, un software de gran potencia gratuito que soporta diferentes lenguajes de programación como C++, C#, Visual Basic .NET, Java, Python, Ruby y PHP. Visual es una plataforma creativa que se utiliza para editar, depurar y construir código. En este caso, se utilizó lenguaje C# para la creación de los comportamientos que deben tener cada uno de los GameObjects en las Escenas y por tanto, el funcionamiento de la aplicación en general.

<sup>52</sup> <https://docs.microsoft.com/en-us/visualstudio/get-started/visual-studio-ide?view=vs-2019>

C# es un lenguaje de programación desarrollado por Microsoft para su plataforma .NET. Es un lenguaje estructurado, orientado a objetos (OOP). Lo que significa que encapsula atributos y comportamientos en objetos, los cuales pueden ser invisibles entre sí pero guardan una comunicación entre ellos mediante interfaces definidas. La diferencia con lenguajes procedimentales como C, deriva en que estos están orientados a acciones mientras que C# está orientado a objetos. A las funciones en C# se les conoce como métodos y trabajar con clases y métodos simplifica bastante la codificación de un programa, ya que permite la reutilización de códigos y la formación de nuevas clases heredando atributos y comportamientos de otras ya existentes. La herencia y el polimorfismo componen la base del lenguaje orientado a objetos. Polimorfismo se refiere a la posibilidad de escribir códigos de forma general con el fin de administrar varias clases interrelacionadas (Deitel & Deitel, 1998). El poder de C# se deriva de C/C++ e incluye mejoras de JAVA y .NET. Su sintaxis hace que sea de fácil aprendizaje y aplicación puesto que permite una arquitectura vertical eficaz en cuanto a expansión multiplataformas.

### **2.2.3.3 SDK de Audio 3D y Ambisonics**

Los SDK (Software Development Kit) son paquetes de herramientas que sirven a los desarrolladores para facilitar la creación de aplicaciones en distintas plataformas, lenguajes y sistemas operativos. En estos kits se encuentra la documentación del paquete, el API<sup>53</sup>, proyectos de ejemplo, librerías, controladores y protocolos. La mayoría de SDK son gratuitos. Los fabricantes de dispositivos suelen ofrecer sus paquetes de código abierto para que los desarrolladores realicen aplicaciones para sus productos.

Es relevante mencionar que para el presente desarrollo se utilizaron al menos dos SDK. Uno de ellos es el SDK de Oculus para desarrollar aplicaciones de realidad virtual. En la versión de Unity 3D 2020, es necesario contar con los paquetes de XR Interaction Toolkit<sup>54</sup>, XR Interaction Subsystem, XR Plugin management y el Oculus XR plugin<sup>55</sup>. Estos han apuntado a unificar las herramientas de desarrollo para todos los dispositivos de Realidad Virtual que hay en el mercado, para de esta forma no tener que utilizar un SDK por fabricante

---

<sup>53</sup> Application Programming Interface. Es una interfaz que define las interacciones entre el software y múltiples intermediarios.

<sup>54</sup> <https://docs.unity3d.com/Packages/com.unity.xr.interaction.toolkit@0.9/manual/index.html>

<sup>55</sup> <https://docs.unity3d.com/Packages/com.unity.xr.oculus@1.4/manual/index.html>

sino que de un solo paquete se pueda desarrollar para varios dispositivos<sup>56</sup>. Por otra parte, se cuenta con dos opciones de SDK para audio espacial, los cuales ofrecen distintas funcionalidades. El SDK multiplataforma de Google se conoce como Resonance Audio<sup>57</sup>. Este ofrece a los desarrolladores y diseñadores sonoros una herramienta poderosa con tecnología de alta fidelidad enfocada para dispositivos móviles con límites en recursos computacionales. Del otro lado está el SDK de HTC llamado VIVE 3DSP Audio SDK<sup>58</sup>. Este kit es una solución para simular sonido espacial realista en un entorno virtual. El SDK puede ser utilizado para crear audio 3D atractivo e inmersivo para aplicaciones básicas o para experiencias en Realidad Virtual. Este SDK tiene ciertas funcionalidades que lo hacen superior al Resonance. Desde el punto de vista de la documentación, contiene mucha más información detallada que permite hacer modificaciones exactas en los scripts<sup>59</sup>. También permite trabajar con audio de alta resolución a 48.000Hz. Provee HRTFs precisas con ángulos de variación de 5 grados, Ambisonic de tercer orden, efectos por oclusión y una de las funciones más interesantes, permite guardar el audio espacial en un fichero WAV.

Ambisonics<sup>60</sup> es una tecnología de sonido envolvente la cual utiliza una esfera completa alrededor del oyente para simular el sonido espacial. Los armónicos en la esfera se presentan a medida que la presión del sonido cambia alrededor de la cabeza del oyente. El orden del Ambisonic determina el nivel de precisión y realismo en la dirección de los sonidos. Por lo cual, un tercer orden en los armónicos de la esfera genera una mejor directividad de los sonidos en el espacio. Entre mayor es el orden, mayor es el consumo computacional. Sin embargo el SDK de HTC tiene una patente que permite trabajar en tercer orden con el consumo de un Ambisonic de primer orden. En la figura 18 se muestra una representación de lo que hace la tecnología Ambisonic. La imagen izquierda análogamente muestra la esfera como si tuviera altavoces en puntos distintos. Y la figura de la derecha muestra una representación del orden de los armónicos esféricos.

---

<sup>56</sup> Actualmente hay soporte para todas las gafas de Oculus, pero aún no se incluyen las HTC por lo cual es necesario utilizar el SDK de Steam.

<sup>57</sup> <https://resonance-audio.github.io/resonance-audio/>

<sup>58</sup> [https://hub.vive.com/storage/3dsp/vive\\_3dsp\\_audio\\_sdk.html](https://hub.vive.com/storage/3dsp/vive_3dsp_audio_sdk.html)

<sup>59</sup> Los Scripts son los códigos de programación que controlan el comportamiento de los objetos.

<sup>60</sup> <http://www.ambisonic.net/>

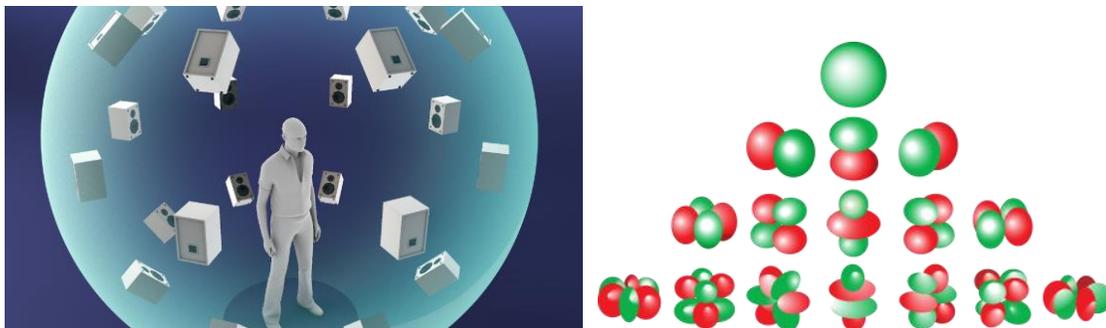


Figura 18 Representación del Ambisonic y el orden de los armónicos

Fuente: izquierda obtenida de <https://resonance-audio.github.io/resonance-audio/discover/concepts.html> y derecha obtenida de [https://hub.vive.com/storage/3dsp/vive\\_3dsp\\_audio\\_sdk.html](https://hub.vive.com/storage/3dsp/vive_3dsp_audio_sdk.html)

## 2.2.4 Mezclador Visual

### 2.2.4.1 Imaginación

Como se trató de explicar en el apartado sobre sonido y señales, existen al menos dos procesos que influyen en la transmisión e interpretación de los sonidos. En adelante, cuando se hable de sonidos se tratará de música, a no ser que se indique lo contrario. El primer proceso relacionado con la transmisión del sonido se da mediante un fenómeno físico, en donde las ondas acústicas viajan por un medio. Cuando se escucha música, lo normal es que estas ondas sean creadas por el movimiento de los conos de los parlantes en los altavoces, los cuales desplazan las partículas del aire en todas las direcciones dentro del cuarto o el lugar donde se encuentre el oyente. Las ondas golpean cada superficie y también cada célula del cuerpo humano que las percibe. De allí que también sea posible decir que la música no solo se escucha sino que se siente. Lo anterior abre la discusión hacia otro proceso necesario para la percepción, el cual es conocido como la imaginación. En el apartado sobre el funcionamiento del sistema auditivo se clasificó la imaginación como la etapa 5, relacionada con la decodificación que hace el cerebro de los estímulos recibidos desde el oído interno. La imaginación juega un papel relevante en los procesos de mezcla de una canción y en general en la percepción que tienen las personas de una pieza musical. Sin los procesos de imaginación no sería posible imaginar el comportamiento de los sonidos que salen de los parlantes, como cuando se está dormido. En realidad, el oído y el cuerpo siempre están sintiendo las ondas acústicas, pero si no hay un proceso de atención consiente sobre lo que se escucha no es posible imaginar los sonidos y por tanto analizarlos. Lo mismo sucede cuando se escucha música por medio de audífonos, se percibe como si los sonidos estuvieran

en medio de la cabeza (o en el interior), pero en realidad allí no hay más que un cerebro (Gibson, *The Art Of Mixing A Visual Guide To Recording Engineering And Production.* , 1997).

La imaginación hace posible ver más allá de los parlantes. Y en muchas ocasiones los músicos, productores o ingenieros de mezcla cuando están organizando los sonidos en una canción, miran a un lado u otro para encontrar la ubicación de éstos. Por lo general, la música se basa en técnicas de reproducción estéreo, lo que significa que toda canción tiene dos canales con distinta información, uno para el altavoz derecho y otro para el izquierdo. ¿Cómo es posible que por medio de dos simples altavoces se pueda crear la sensación de movimiento horizontal o vertical de un sonido? Cuando se habló sobre la percepción que tienen las fuentes sonoras en el oyente se indicó que de acuerdo con el lugar de la fuente un oído percibe más o menos sonido directo, en una dirección, intensidad o tiempo y que todo esto es filtrado por el PA para transferir la información de espacialidad hacia el oído medio e interno y por último el cerebro. Pero ahora se trata de un proceso contrario. La imaginación de los sonidos dentro del espacio permitido por los altavoces funciona como un espacio en el cual se organizan los sonidos para crear percepciones, sensaciones y movimiento. En donde el oyente es el encargado de descifrar el arte que hay detrás, y lo que esa combinación le genera en términos de sensación y sentimiento. La reproducción y procesamiento de sonidos grabados da los efectos de distancia y espacio que permiten a la mente reconstruir sus causas. Por último queda decir que el proceso de imaginar los sonidos juega un rol importante en la forma en la cual ingenieros y productores crean piezas musicales. Sin embargo este concepto no es del todo aceptado puesto que en ocasiones es afectado por factores psicológicos del oyente o por elaboraciones predominantes del cerebro en cuanto a estímulos o la falta de estos, haciendo que el proceso imaginativo sea un objeto de estudio subjetivo (Agostini, 2014).

#### **2.2.4.2 Estéreo y mono**

En medio de dos parlantes puede no haber físicamente nada, puede haber una ventana o una pared. Pero aun así, hay sonidos que se pueden imaginar en el centro de los parlantes o en los costados. Esto en los procesos de mezcla se conoce como mid-side<sup>61</sup>. Donde mid, es

---

<sup>61</sup> También se conoce como Campo Estéreo y define el espacio virtual el cual solo existe en la mente del oyente. Allí se posicionan los sonidos individuales o el conjunto de una mezcla.

un acrónimo de middle o centro, y es el lugar en donde se ubican los sonidos principales de una canción. Mientras que side, está relacionado con los costados y sirve para crear la sensación de espacialidad, de una mezcla abierta cuyos sonidos son percibidos como si salieran más hacia delante o arriba de los altavoces. De lo anterior, se puede inferir que la diferencia entre un sonido mono y uno estéreo, radica principalmente en que la información de intensidad, fase y frecuencia es la misma en ambos canales para el sonido mono, mientras que en estéreo es distinta. Se puede tener un sonido mono de un solo canal o de dos, pero no se puede tener un sonido estéreo con un solo canal.

Por otra parte, la reproducción estéreo conlleva a que se presenten ciertos problemas en la percepción y por tanto, en la imaginación de los sonidos. Uno de los problemas se debe a que no es posible controlar la cantidad de señal que un altavoz entrega en el oído opuesto (sin contar con que el oyente puede mover la cabeza)<sup>62</sup>. Otro inconveniente es que rara vez se puede percibir la música en un entorno acústicamente tratado, por tanto la sala o el recinto influye en menor o mayor medida en la respuesta frecuencial de los altavoces, lo cual cambia la manera en la que el ingeniero de mezcla, productor u oyente imagina los sonidos, lo cual repercute en las decisiones que se toman al momento de organizar una pieza musical. De lo anterior es posible inferir que un ambiente ideal en cierta forma sea la escucha por medio de audífonos, los cuales desconectan el efecto acústico de la sala y logran una mejor imagen estéreo con la salvedad de que se pierda un poco la percepción de espacio en cuanto a la profundidad. De allí que los sistemas holofónicos utilicen audífonos para reproducir el sonido filtrado por la respuesta al impulso de los maniqués usados para obtener las HRTF. Los audífonos serían el entorno ideal de escucha sino fuera por el efecto in-head, el cual hace que los sonidos se perciban dentro de la cabeza y no como si estuvieran en el exterior. Por ello aquí es en donde el audio 3D o sonido espacial, mediante técnicas de holofonía, solucionan ese problema, puesto que por medio de atenuaciones respecto a la distancia logran el efecto de lateralización de las fuentes sonoras.

---

<sup>62</sup> Aunque existen sistemas de cancelación cross-talk usados para la reproducción binaural con altavoces, no es tan común usar esas técnicas en estudios de producción musical.

### **2.2.4.3 La idea tradicional de los controles.**

A lo largo de la historia de los sintetizadores hubo un momento en donde la incursión de lo digital cambió las relaciones entre el músico y el instrumento por medio de la inclusión de pantallas LCD y botones, distanciándose del flujo de trabajo con los faders y los knobs<sup>63</sup> clásicos. En ese entonces los nuevos equipos tenían todo un manual de instrucciones en la carcasa el cual indicaba la función de cada input y output. Esto conllevó a un periodo de frustración, pues para dominar un equipo sintetizador FM por ejemplo, había que dedicar tiempo y paciencia para lograr sonidos nuevos distintos a los preestablecidos de fábrica. En el caso de las consolas de mezcla, este escenario continúa en progreso y a lo largo de los años se ha buscado versatilidad en lo digital en sincronía con el control natural de lo analógico, pero obtener equipos de este nivel resulta en un costo económico alto. Sin embargo, la evolución de la tecnología ha permitido la incursión de nuevas tendencias en el control de los sintetizadores y las consolas incluyendo entradas táctiles multitouch, sensores de luz, de presión, de movimiento, entre otros.

### **2.2.4.4 La consola de mezcla**

En esencia una consola de mezcla es un equipo que permite la entrada de señales de audio por canales, los cuales son sumados y entregados en una salida estéreo. Sin importar la tecnología que se maneje, una consola de mezcla se comporta como un sumador de señales, si esta es analógica, la suma será de valores de voltaje en el tiempo. Si es digital, se realiza una conversión al dominio digital para sumar bits y luego volver a convertir ese resultado en una señal analógica en el dominio del tiempo. Existen diferentes tipos de consolas, algunas varían de acuerdo con los efectos y calidad de los componentes electrónicos que tienen, así también como los diseños internos y funciones. Pero en esencia, el esquema básico de una consola no ha cambiado desde los 50's. Algunas incluyen funciones de almacenamiento de parámetros, de automatizaciones, de proyectos, involucrando lo digital y lo analógico para una mayor libertad de administración de los ruteos de las señales. Algunas incorporan comunicación directa con una computadora y otras funcionan para grabar directamente en la consola por medio de discos duros internos o CD-ROM<sup>64</sup>.

---

<sup>63</sup> Deslizadores y perillas. Son los controles de mando básicos en una consola de mezcla.

<sup>64</sup> Como la estación digital de audio Yamaha AW4416.

La mayoría de los mezcladores utilizan componentes electrónicos similares, como faders y perillas para controlar niveles de volumen o aplicar efectos. Desde el análisis de circuitos tanto los faders como los knobs (o perillas), son potenciómetros. Éstos varían el voltaje de entrada de acuerdo con la resistencia que presente el potenciómetro, lo cual se conoce como un divisor de voltaje. La diferencia está en la forma y los valores de voltajes y resistencia que presenta uno y otro. Sin embargo, existen otros tipos de knobs que son ampliamente usados en consolas y que no solo son potenciómetros, sino que pueden tener valores digitales, como los Encoders o pueden ser de alta precisión como los potenciómetros logarítmicos.

Por medio de los botones, perillas y faders, los mezcladores tienen funciones básicas que se presentan en conjuntos. A esto se le conoce como Channel Strip, y son funciones que se repiten en cada canal del mixer en donde se encuentran la ganancia de preamplificación de la señal de entrada, un ecualizador como mínimo con tres bandas de ajuste de la señal luego de ser amplificada, seguido por una parte de panning estéreo donde una perilla establece el valor de la señal para los canales derecho e izquierdo. Luego continúa una parte de envíos auxiliares de la señal hacia otros canales o salidas con efectos, ya sean internos o externos a la consola. Por último, la parte inferior usualmente tiene botones de Mute y Solo del canal, los cuales sirven para silenciar ese canal o para silenciar los demás canales. Al final está el fader de volumen, el cual regula la cantidad de voltaje que ese canal entregará a la suma total del canal Master<sup>65</sup>. El canal master es el encargado de sumar las señales de los canales de entrada y usualmente también cuenta con un las funciones de ecualización, panning, envíos y volumen. También hay salidas de monitoreo por audífonos o salidas secundarias para otros tipos de seguimiento de las señales o de los envíos auxiliares. Algunas consolas más avanzadas tienen medidores digitales, analizadores de espectro en tiempo real, y pueden agrupar canales de entrada en canales individuales para manipular varias señales en un solo Channel Strip. También algunas consolas incluyen compresores dinámicos por canal individual o en el master.

Actualmente existen consolas de varios precios y marcas, con nuevas funciones y mejoras. Y aunque la evolución de las tecnologías digitales ha hecho reducir el uso de las consolas

---

<sup>65</sup> A todo bus final en una consola, ya sea digital o analógica, hardware o software, se le conoce como canal Master y es el encargado de realizar la suma final de todos los canales.

aún siguen siendo el mejor complemento de trabajo en estudios profesionales y estudios caseros. En la figura 19, se puede observar a Paul McCartney con una consola icónica de los estudios Abbey Road<sup>66</sup>.



*Figura 19 Paul McCartney con la consola TG12345<sup>67</sup>*

*Fuente: obtenido de <https://www.abbeyroad.com/news/behind-abbey-road-studios-emi-tg12345-console-2604>*

#### **2.2.4.5 Estación de trabajo de audio digital (DAW)**

También conocido como DAW (por sus siglas en inglés Digital Audio Workstation) es la primera gran revolución del concepto tradicional de las consolas de mezcla. Y más que una revolución, es una evolución gigante que no solo está relacionada con las consolas sino con todo el estudio de grabación. Tanto así, que gracias a este avance existen los estudios caseros en la forma actual y de igual manera, los estándares contemporáneos en la producción musical desde los grandes estudios. Un DAW es un software que permite la grabación y edición no destructiva de señales de audio. Esto último es uno de los principales aspectos que marcó la diferencia, puesto que anteriormente las grabaciones se realizaban en cintas y en equipos analógicos en donde las ediciones eran destructivas, es decir, una vez editado o grabado el

---

<sup>66</sup> Se recomienda ver un sitio creado por Google llamado Inside Abbey Road en donde por medio de un paseo interactivo se puede ingresar a uno de los estudios de grabación más icónicos y famosos de la historia. <https://www.abbeyroad.com/news/google-presents-inside-abbey-road-1003?AspxAutoDetectCookieSupport=1> o también en este enlace <https://www.stinkstudios.com/work/inside-abbey-road>

<sup>67</sup> Se recomienda ver Chaos and Creation At Abbey Road de Paul McCartney. Allí se muestran los equipos con los cuales se realizaron sus mejores canciones y claro, un concierto impecable. <https://www.youtube.com/watch?v=9eIQeVfrLOo>

audio, no era posible volver a un punto anterior, lo cual implicaba volver a grabar en el caso de haber cometido un error en la edición. Con la llegada de los DAW, la consola de mezcla y los procesos de edición de audio pasaron al ámbito digital, en donde ya no era necesarias las cintas sino los discos duros, ya no eran necesarios los equipos costosos y de gran tamaño porque estos se podían incluir en el flujo de trabajo dentro de los DAW. Muchas de las interfaces gráficas de usuario de los DAW están pensadas para guardar relación con los anteriores equipos de hardware, lo cual mantienen una fuerte conexión entre el pasado y el presente. Los DAW permiten análisis espectrales en tiempo real, por ejemplo, la interfaz de un ecualizador permite mostrar la función de transferencia de un filtro sin afectar la señal de entrada, permitiendo su alteración sin dañar el archivo de audio original. Un compresor permite mostrar la cantidad de dB que reduce en tiempo real, al igual que un limitador. La mayoría de efectos que se pueden encontrar en los DAW tienen sus propias interfaces gráficas y parámetros para controlar las señales. Estos efectos pueden ser desarrollados por terceros y gracias a una tecnología llamada VST (Virtual Studio Technology)<sup>68</sup> es posible integrar estos software en forma de sintetizadores y efectos.

Los DAW no solo funcionan para técnicas estereofónicas sino para sonido envolvente 5.1,7.1, sino también para sonido ambisónico. La mayoría de estos programas incorporan una interfaz gráfica que mínimo tiene dos partes, una en donde se ubican los audios en una línea de tiempo y otra en donde se extrapola el concepto de Channel Strip de las consolas de mezcla, a esto normalmente se le conoce como secuenciador y mixer. Actualmente existen una gran variedad de DAW los cuales ofrecen distintas funciones y enfoques. El estándar de la industria musical es Pro Tools<sup>69</sup> y ha sido catalogado así, porque la mayoría de las grandes producciones musicales se han hecho en ese software. Sin embargo existen muchos otros más de distintos precios y características. Por ejemplo, el DAW de Steinberg para sonido espacial es Nuendo<sup>70</sup> y uno de los mejores DAW para Mastering es Sequoia<sup>71</sup>. En la figura 19 se muestra la interfaz gráfica de uno de los DAW más utilizados en la industria junto con algunos de sus instrumentos virtuales.

---

<sup>68</sup> Inventada por Steinberg <https://developer.steinberg.help/display/VST/VST+Home>

<sup>69</sup> <https://www.avid.com/pro-tools>

<sup>70</sup> <https://new.steinberg.net/nuendo/>

<sup>71</sup> <https://www.magix.com/us/music/sequoia/>



Figura 20 Interfaz gráfica de Pro Tools

Fuente: obtenido de <http://www.avidblogs.com/pro-tools-11-virtual-instruments-for-music-and-audio/>

#### 2.2.4.6 Conceptos básicos de un mezclador visual

Con la información detallada hasta el momento, es posible tener un marco de referencia sobre el cual plantear conceptos funcionales para el desarrollo de este proyecto. Con base en las definiciones anteriores, en la conceptualización de las herramientas básicas de un estudio, su utilización y efectos, se debe preparar la idea de una herramienta para la visualización de los sonidos dentro de una aplicación de Realidad Virtual. En lo que respecta al porqué del término visual, es posible argumentar que las interfaces gráficas son necesarias porque conllevan herramientas ya existentes a la mente, en donde estas se transforman en elementos objetivos e inequívocos. Por otra parte, los ingenieros y técnicos de audio, en su flujo de trabajo usan herramientas visuales a menudo, como se logró definir en el apartado que trata sobre los DAW. De esto último, se infiere sobre la necesidad de herramientas visuales que permitan un mayor tiempo de trabajo efectivo sin fatiga auditiva ni cansancio mental. Una herramienta de recalibración auditiva podría ser esencial para conseguir mejores resultados en proyectos que pueden ser frustrantes, en el sentido en que pueden tomar bastante tiempo en su finalización. Por tanto, cabe preguntarse sobre lo que debe ofrecer una herramienta de mezcla visual. Específicamente para sonido espacial. En ese sentido, y basado en (Agostini, 2014) y (Gibson, The Art Of Mixing A Visual Guide To Recording Engineering And Production. , 1997) es posible afirmar que una interfaz de este tipo debe considerar de manera

práctica e intuitiva el control de los parámetros relevantes a los sonidos. Para el control de estos parámetros será pertinente un acceso inmediato y versátil que permita una manipulación fluida, alejándose de las representaciones artísticas que puedan desviar el enfoque del desarrollo<sup>72</sup>. Por tanto, el mezclador visual deberá proporcionar una neutralidad en el estilo y necesidades del usuario, tal cual como lo han hecho las consolas de mezcla tradicionales. Al mismo tiempo debe prestar las opciones necesarias para poder localizar y modificar la espacialidad de los sonidos individualmente.

#### **2.2.4.7 El arte de mezclar**

David Gibson en su libro sobre mezcla trata el problema de la representación de los sonidos desde una postura totalmente nueva. Plantea un modelo orientado hacia la visualización de la imagen de los sonidos de manera didáctica, en donde busca que el oyente imagine el movimiento y características de cada sonido, así mismo como los efectos que se pueden aplicar en los procesos de mezcla de una canción. El libro cuenta con un DVD y un sitio web<sup>73</sup> en donde hay más información sobre cursos y otras investigaciones, sin embargo no se ha encontrado una muestra comercial de este sistema de mezcla<sup>74</sup>.

Uno de los mayores aportes de Gibson, radica en la orientación pedagógica y la objetivación de los sonidos para discutir técnicas básicas y principios de mezcla. La comunicación efectiva de conceptos y enunciados difíciles de entender hacen que su obra sirva de introducción al mundo de la producción musical de una forma no vista con anterioridad. A este respecto, lo que Gibson trae al lector es una transformación, una reconstrucción de un conocimiento técnico específico reducido en expresiones simples, de fácil aprehensión y aplicación. Lo que significa una traducción de los conocimientos científico-tecnológicos a un contexto cotidiano, lo que en otras palabras se conoce como transposición didáctica (Vázquez Alonzo & Alarcón Zamora, 2010). Finalmente, el aporte de Gibson se puede tomar como el rompimiento de la barrera entre las experiencias y conocimientos del

---

<sup>72</sup> Sobre esto cabe resaltar que se busca una forma de representar los sonidos de manera simple y concisa sin la necesidad de involucrar diseño gráfico que conlleve a un trabajo relacionado con el arte en sí de cada sonido, pues sería toda una tesis adicional.

<sup>73</sup> <https://soundhealingcenter.com/love/vm/>

<sup>74</sup> Aunque no existan muestras comerciales del aporte de Gibson, investigando un poco se encontró la patente que establece la representación visual de los sonidos como un invento suyo (United States Patente nº US 6,490,359 B1, 2002)

profesional que no pueden ser comunicados mediante el lenguaje y que basado en lo visual, encuentra un transductor de gran eficiencia para transmitir sus ideas.

### 3 Metodología

El desarrollo de software es un proceso que en la mayoría de los casos involucra la construcción de un producto o de un prototipo. De allí que se analice el ciclo de vida de un desarrollo como el tiempo desde el cual una propuesta es conceptualizada, implementada, entregada, puesta en marcha y documentada. En ese sentido, para obtener buenos resultados la mejor forma de trabajo se consigue mediante la incorporación de procesos efectivos y ágiles, que permitan satisfacer los requerimientos solicitados sin desaprovechar recursos ni tiempo (Pfleeger & Atlee, 1998). Con base en lo anterior, el ciclo de vida del proyecto realizado se describe en los siguientes procesos:

- Análisis de requisitos (objetivos).
- Diseño de los escenarios (realización de varios prototipos).
- Diseño de las interfaces de usuario (implementación de varias técnicas).
- Escritura de los códigos de programación (mecánicas).
- Pruebas de validación y verificación (ciclos de repeticiones).
- Pruebas generales de todo el sistema.
- Documentación del proyecto.
- Entrega del prototipo final.

Para el desarrollo de este proyecto se utilizó el Modelo de Prototipo el cual hace parte de los Modelos de Procesos utilizados en desarrollo de software, en donde básicamente se tiene un sistema de requerimientos como entrada y una entrega de producto final como salida (ver figura 21). Los requerimientos fueron los objetivos planteados con el problema de investigación, y cada uno de los objetivos específicos fue un requisito que llevó al cumplimiento total del objetivo general.



Figura 21 Esquema Básico de los modelos de procesos

Fuente: elaboración propia

### 3.1 Modelo de prototipo

El modelo permitió que las partes del sistema a desarrollar fueran construidas de manera rápida y eficaz para entender y aclarar dudas sobre los requerimientos del proyecto. Lo que implicó una investigación repetida de los requisitos de diseño para asegurar una comprensión estricta de lo que se necesitaba y lo que se propuso.

Los ciclos de repetición de este modelo pueden ser agregados o eliminados dependiendo de la necesidad del prototipo. Por lo tanto, para el desarrollo de la aplicación se necesitaron ciclos de repetición en todas las etapas de los procesos, puesto que si bien al principio se tenía claro el objetivo a cumplir, existió bastante incertidumbre frente a varios aspectos estructurales. Lo anterior motivó la elección de este modelo como el más apropiado para administrar las etapas de la investigación, puesto que el modelo de prototipo reducía el riesgo y las incertidumbres en el desarrollo, permitiendo contrastar los requerimientos que en realidad debería tener el producto final. En la figura 22 se muestran las etapas del modelo a nivel general, en donde se puede observar que el modelo ofreció ciclos de repetición constante, los cuales permitieron reformular o apoyar los requerimientos inicialmente solicitados.

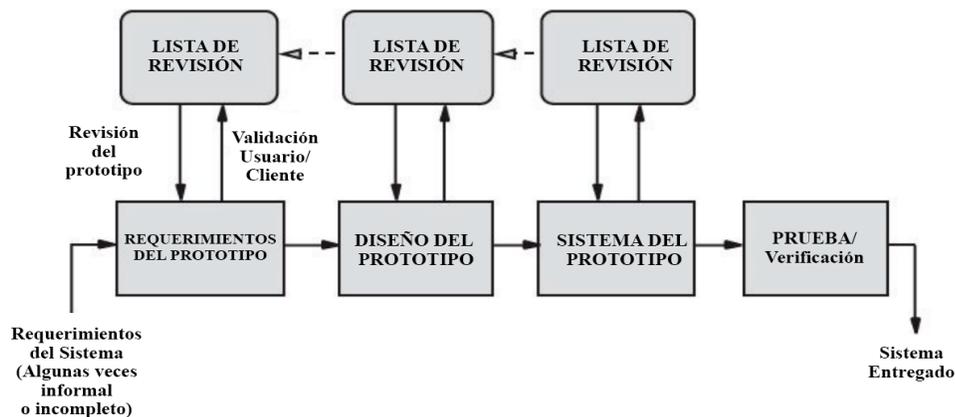


Figura 22 Modelo de Prototipo

Fuente: modificado de (Pfleeger & Atlee, 1998)

La permanente revisión de los requisitos permitió que una vez definido el diseño para los escenarios y las interfaces gráficas de usuario, se diera paso a la construcción de prototipos y por ende a la programación de sus mecánicas. En la siguiente figura se puede validar el ciclo de vida del proyecto, junto con los ciclos de repetición.

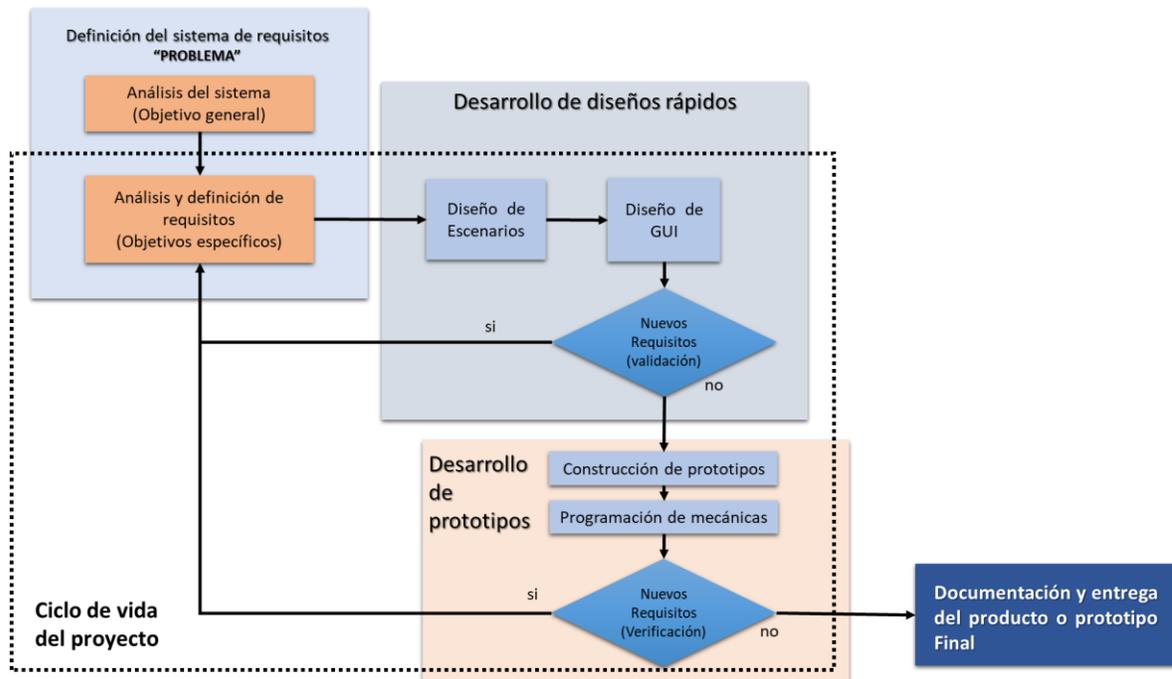


Figura 23 Ciclo de vida del modelo de prototipo

Fuente: modificado de <https://gestionrrhhusm.blogspot.com/2011/05/modelo-de-prototipo.html>

### 3.2 Etapas metodológicas

La aplicación se dividió en cuatro procesos generales. En el primer proceso se analizó el sistema de requisitos que se debían cumplir. En el segundo proceso se diseñaron los ambientes virtuales que permitieron una mejor experiencia inmersiva y también se implementaron las interfaces gráficas para los diferentes menús. En el tercer proceso se realizó la programación de las funciones de los sonidos y de las interfaces, así también como las mecánicas de interacción entre el usuario y la aplicación. Finalmente, luego de haber aprobado las validaciones y verificaciones de cada ciclo, se documentó el proyecto para ser entregado. A continuación se describen estos procesos de manera detallada.

El primer proceso que se abordó estuvo relacionado con la definición del sistema de requisitos de la investigación, es decir, con los objetivos, alcances y delimitaciones. Lo cual implicó directamente una detallada revisión del marco teórico y por tanto, de los antecedentes. En la medida en que se obtenía más información, más se ampliaba el horizonte y las posibilidades que ofrecía investigar sobre audio 3D y por consiguiente, se moldeaba la forma estructural del planteamiento del problema. En ese sentido, la mayoría de referencias y antecedentes se encontraron en publicaciones y universidades del exterior, como el caso de

la tesis de maestría de (Agostini, 2014), la cual fue enviada directamente desde la Universidad de Roma. Para encontrar esa referencia fundamental en la presente investigación, se tuvo que realizar una búsqueda intensiva en YouTube y en otros buscadores web. En la descripción del video de Agostini<sup>75</sup>, se encontró la referencia de la tesis y de la universidad, pero al no encontrar el enlace del documento se tuvo que contactar formalmente a la facultad para solicitar el archivo. Del análisis riguroso de ese documento, se encontraron datos importantes sobre las patentes de Gibson y referencias a otros trabajos de especialistas en el campo de la ingeniería de sonido, lo cual posicionó de manera definitiva el rumbo del desarrollo elaborado. Por otra parte, las recientes actualizaciones en los paquetes de programación para Realidad Virtual de Oculus y Unity involucraron un trabajo de estudio considerable, ya que los SDK y plugins cambiaron sustancialmente en el último año imponiendo un tratamiento completamente nuevo y por tanto, era relevante identificar las herramientas necesarias para iniciar el proyecto. En esta primera etapa metodológica, también se tuvieron en cuenta los alcances y delimitaciones que debía tener la aplicación. Tales límites fueron marcados por la extensión de lo investigado en el marco teórico y las formulaciones propuestas desde la base de los antecedentes.

El segundo proceso estuvo relacionado con la creación de ambientes y escenarios. Por lo tanto, para la obtención del entorno principal se realizaron dos prototipos, los cuales fueron evaluados en términos de navegación, escala e inmersión. Si bien el primer prototipo fue una idea que permitió tener claridad sobre el nivel inmersivo que se buscaba en cuanto la gamificación del desarrollo, éste no cumplió con los parámetros de navegación ni de escala que se buscaban, ya que no tenía en cuenta las afectaciones por Motion Sickness y el usuario era más pequeño en relación con los objetos de ambientación. Ese primer intento dio paso al segundo prototipo, en el cual se implementó la técnica Teleport para evitar afectaciones en la salud del usuario y también se definió el diseño gráfico de los principales escenarios, los cuales se realizaron a escala uno a uno con respecto a las dimensiones de los objetos reales. Lo anterior también sucedió con los menús e interfaces gráficas de usuario, por ejemplo en la parte del tutorial. Inicialmente se desarrolló un prototipo de tutorial por ejemplos y acciones básicas del usuario con objetos de prueba. Eso no resultó ser funcional, y por medio

---

<sup>75</sup> <https://www.youtube.com/watch?v=vxoxCpYCE-c>

de pruebas con personas sin conocimientos previos de la aplicación se logró identificar que los usuarios no lograban interactuar de la mejor forma, por lo que se decidió implementar otro sistema haciendo uso de videos, donde se explicaron los botones que el usuario debía presionar para realizar una acción dentro de la aplicación, lo cual fue un prototipo exitoso que funcionó muy bien y superó las pruebas de validación y de verificación.

El tercer proceso investigativo se llevó a cabo durante la programación de las funciones de las interfaces y de las mecánicas de interacción con los sonidos. Para encontrar los SDK adecuados, se tuvo que buscar rigurosamente en foros y tutoriales, en las herramientas de ayuda y en los espacios de aprendizaje que ofrece Unity. De esta búsqueda en foros se encontró el paquete de HTC 3DSP, el cual permitió tener acceso libre al código fuente para realizar modificaciones que permitieran por ejemplo la grabación del audio, el cambio de reverberación de la sala por medio del cambio de los materiales de las paredes, pisos y techos, la aplicación de filtros frecuenciales a los sonidos, entre otros. Por otra parte, durante toda la programación de los diferentes códigos, se utilizó el manual de Unity, los ejemplos y los tutoriales de Realidad Virtual y de los componentes de Audio. En la codificación se invirtió la mayor cantidad de tiempo del proyecto, porque en algunos casos se tuvo que realizar mínimo hasta 3 algoritmos diferentes para resolver una problemática específica o función.

Por último, en la etapa de formalización y escritura se realizaron al menos seis versiones, en las cuales se sugirieron correcciones de forma y de fondo e incluso, la creación del documento permitió alcanzar una mayor claridad sobre el desarrollo y definió los últimos detalles. Ese ejercicio de plasmar lo elaborado durante el tiempo de investigación y de trabajo riguroso, permitió profundizar bastante en la comprensión de la naturaleza del sonido, en la evolución y conceptualización de los dispositivos de Realidad Virtual y por ende, en la constatación de los resultados con respecto a los objetivos propuestos, avalados por el marco teórico.

## **4 Desarrollo**

A continuación se relaciona de manera detallada las partes fundamentales del prototipo desarrollado. En el marco teórico se logró delimitar conceptualmente los parámetros mínimos que debía tener la aplicación para trabajar con audio espacial y permitir al usuario una experiencia inmersiva en Realidad Virtual. Por tanto, las decisiones que se tomaron para el desarrollo estuvieron sujetas a la teoría y a la verificación constante del funcionamiento de todas las opciones, en sincronía con el manejo adecuado de Unity.

### **4.1 Herramientas utilizadas**

La aplicación se desarrolló en el motor de videojuegos de Unity 3D versión 2020, en conjunto con Visual Studio 2019. El lenguaje de programación utilizado fue C# y el SDK implementado para sonido espacial fue el de la empresa HTC VIVE 3DSP Audio SDK. El proyecto fue realizado con las gafas de Realidad Virtual Oculus Quest para la plataforma de Oculus PC, lo cual permitió que la aplicación sea funcional en dispositivos anteriores como las gafas Oculus Rift, Rift S y HTC VIVE por medio de la herramienta Revive. Para la conexión entre las Quest y la computadora se utilizó el accesorio Oculus Link y se utilizaron texturas y modelos 3D gratuitos, junto con algunos Assets que fueron de pago y otros que fueron construidos manualmente.

### **4.2 Características de los Recursos**

Para el desarrollo de la aplicación se utilizaron las siguientes herramientas de trabajo.

- Computador Acer Nitro 5. Procesador Core i5 de 8va generación. Memoria RAM de 16GB. Tarjeta gráfica Nvidia GTX 1050Ti de 4GB. Disco duro M.2 de 512GB y HDD de 2TB interno.
- Gafas Oculus Quest.
- Cable USB 3.0 marca Anker Cable (para conectar las gafas al computador).

### **4.3 Assets y scripts de terceros**

El desarrollo de los escenarios involucró la utilización de modelos 3D relacionados con equipos de audio y diseño arquitectónico. Algunos de estos modelos y assets fueron obtenidos gratuitamente mientras que otros fueron construidos manualmente con la

herramienta de Unity llamada Probuilder. En el anexo 4 se encuentra el listado de Assets utilizados con información relevante. Sin embargo, a continuación se describen los Assets más importantes en el desarrollo del proyecto y cómo éstos fueron implementados.

En Unity Learn se encuentran paquetes con proyectos que sirven para el aprendizaje de funciones específicas del motor. Uno de esos paquetes es el llamado “VR Beginner: The Escape Room”, el cual se puede utilizar en sincronía con el tutorial dispuesto por Unity para aprender a usar el nuevo SDK para Oculus XR. VR Beginner cuenta con una plantilla idónea para comenzar a desarrollar proyectos para Realidad Virtual, ya que el paquete de integración para XR cuenta con todo un sistema de interacciones que permite el uso no solo de dispositivos de Oculus sino también de otras marcas. Entonces, para este desarrollo se utilizó la plantilla que provee el paquete de VR Beginner, la cual incluyó los Assets para la técnica de desplazamiento Teleport, junto con los círculos y las animaciones en los Shaders. Las funciones de Teleport ya vienen incluidas dentro del SDK de Unity Oculus XR. Para su implementación se utilizó el script llamado “Teleportation Anchor”, el cual necesitó el paquete de System Interaction y el sistema de capas de Unity para determinar si el rayo que lo apuntaba está habilitado para interactuar como un Teleport. La verificación entre capas y rayos se hizo mediante el sistema de colisiones de Unity, el cual permitió identificar cuando un objeto colisionaba con otro, o cuando un rayo colisionaba con un objeto y si estos estaban en la misma capa o Layer. Por lo cual, el script del Teleport contuvo funciones de eventos que sirvieron para cambiar comportamientos en otros objetos. Algo para tener en cuenta, es que el rayo que permitió el desplazamiento no fue el mismo para la selección de los menús. El rayo del Teleport se activaba al presionar la palanca tipo joystick hacia adelante en cualquiera de los dos controles. Cuando esa acción era realizada, aparecía un rayo doblado hacia el piso de color verde (ver figura 24). Ese rayo debía ser apuntado hacia abajo y una vez que se iluminaban los círculos, se podía soltar la palanca y automáticamente se realizaba el cambio de posición o la teletransportación, puesto que tanto los círculos como el rayo estaban en una misma capa.



*Figura 24 Rayo Teleport*

*Fuente: elaboración propia*

Adicionalmente, con base en las escenas demo de VR Beginner se descubrió la idea de utilizar guías de usuario dentro de los escenarios, las cuales contienen un diseño alusivo a alguna operación que el usuario pueda realizar para conseguir algo, ya sea desplazarse o mover los objetos. Las guías elaboradas se pueden encontrar en los anexos 2 y 3 en la parte final del presente documento y en la aplicación se encuentran en el escenario del cuarto de control y en el cuarto de mezcla.



*Figura 25 Diseño del menú de mano*

*Fuente: elaboración propia*

Otro Asset importante en el desarrollo de la aplicación, fue el paquete UIKits. Este Asset es de pago y provee a los desarrolladores herramientas UI modificables. Con base en este paquete se implementaron las interfaces de usuario de la aplicación y aunque se realizaron modificaciones de diseño, el uso del sistema de scripts que provee UIKits se utilizó sin

ningún cambio. Por ejemplo, para la creación del menú de mano, se utilizaron diseños propios (ver figura 25) que permitieron modificar las imágenes por defecto del Asset.

El SDK de audio 3D fue pieza clave para el desarrollo del proyecto. Por medio del paquete de HTC Unity Plugin se logró poner en funcionamiento gran parte de las mecánicas relacionadas con las pistas de audio y con la grabación en formato WAV. Para dar al usuario la percepción de audio 3D con respecto al punto de escucha, se utilizó el SDK de HTC Vive 3DSP Audio junto con los componentes básicos de Unity Audio. Para implementarlo bastó con utilizar dos scripts fundamentales en cada sonido. Primero, en Unity todo objeto de audio necesita el componente “Audio Source”, el cual requiere de un “Audio Clip” que es un archivo de audio normal en formato WAV o en MP3. El Audio Source es el componente que administra las funciones de los audios en las escenas, y desde allí se puede indicar a Unity que active los sonidos al iniciar la aplicación, o se pueden activar por código en cualquier momento. También se puede detener o pausar el audio por medio de código o hacer que se reproduzca cíclicamente<sup>76</sup>. Otro script utilizado para espacializar las pistas de audio fue el “3DSP\_AudioSource” del SDK de HTC, el cual se encargó de aplicar los procesos necesarios para generar el efecto 3D. Por medio de las funciones del plugin de HTC se logró grabar los sonidos de la aplicación mediante el almacenamiento del buffer de datos del Audio Listener. El Listener es un componente nativo del sistema de Audio de Unity, comúnmente se agrega al GameObject de la cámara y sirve para reproducir todos los sonidos de la aplicación. Se puede pensar el Audio Listener como los oídos del usuario, sin embargo para este proyecto, el Audio Listener se ubicó en el punto de escucha, dentro del GameObject de la cabeza. De esta manera, implementando modificaciones al script llamado “3DSP\_AudioListener”, se obtuvo la grabación en formato WAV del audio que reproduce el Audio Listener. Los cambios al código permitieron controlar la ubicación de almacenamiento de los archivos, para que cuando se digite el nombre de la grabación en la interfaz del menú “Exportar .WAV”, la información ingresada se transfiera al script de HTC y guarde el archivo con ese nombre en una ubicación indicada. Los scripts del SDK permitieron hacer modificaciones porque son de código abierto.

---

<sup>76</sup> Las funciones del menú móvil de Transporte acceden por código al componente Audio Source para dar play, stop, pause o loop a los sonidos en cualquier momento.

Por otra parte, el sistema de audio de Unity cuenta con un mezclador (ver figura 26) el cual permite manejar los audios individualmente como si fueran canales de una consola de mezcla convencional o canales en un DAW. La diferencia es que en Unity se les conoce como grupos, y estos grupos van a un canal “master” que se conoce como mixer. En Unity se pueden tener varios mixers que combinen una serie de grupos y a su vez, estos mixers pueden derivar en otro mixer. Además, cada efecto nativo de Unity se puede exponer para acceder por código a sus parámetros. El manejo de los mixer por medio de grupos fue la forma más efectiva de administrar los 10 canales del mezclador, ya que permitió controlar el volumen de cada audio mediante código.

Lo anterior fue una de las razones por las cuales no se utilizó el SDK de Google, ya que Resonance Audio no permite trabajar los audios especializados en grupos sino que todos llegan obligatoriamente a un solo canal Mixer.

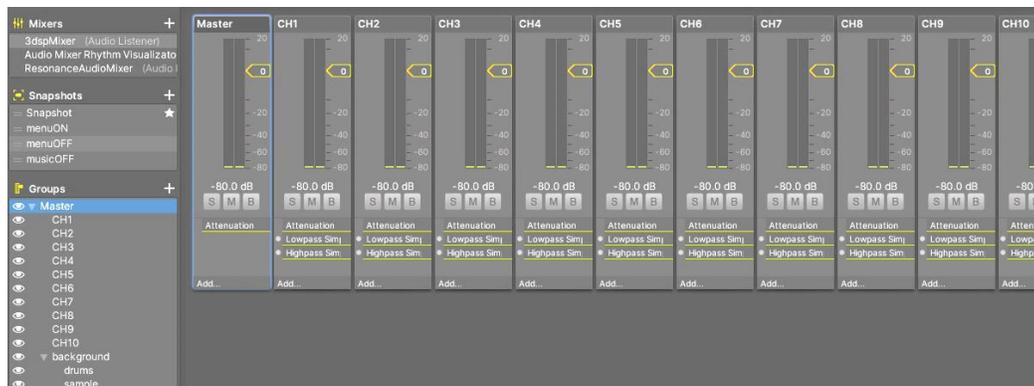


Figura 26 Audio Mixer de Unity

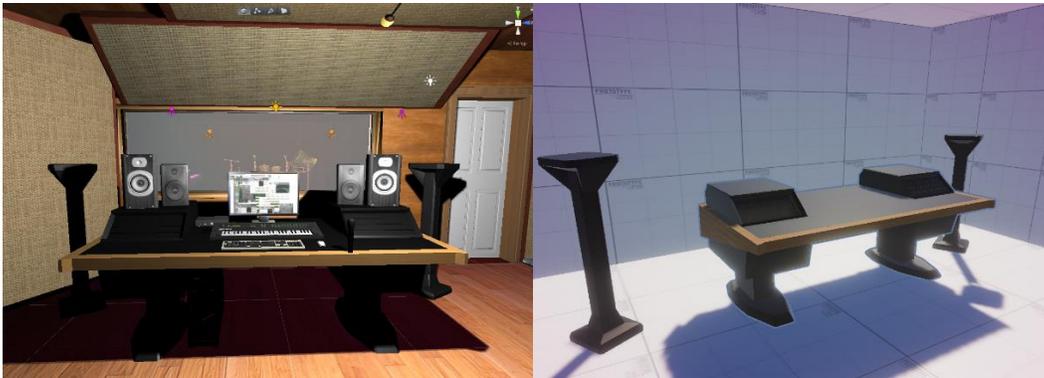
Fuente: elaboración propia

Finalmente otra de las funciones utilizadas del SDK de HTC, fue el cambio de reverberación o la respuesta al impulso del cuarto en donde se aplicó la espacialización de los sonidos. El SDK permitió cambiar el material de las paredes para que las reflexiones tuvieran otro tipo de comportamiento. Para ello, el script llamado “3DSP\_AudioRoom” contó con opciones para activar o desactivar el efecto del cuarto y se logró modificar el tamaño del recinto. El script contó con más de 40 tipos de materiales para las paredes, el techo y el piso. Por lo cual en el menú “Reverb Room”, se utilizó ese script para modificar la respuesta del cuarto de mezcla entre 5 posibles opciones. Esto no se planteó en los

requisitos del proyecto pero fue una opción adicional que funcionó muy bien y se decidió que sería un buen complemento.

#### 4.4 Assets y scripts propios

Los escenarios en su totalidad fueron diseñados a la medida de los requisitos por medio de las herramientas disponibles en Unity. Con el paquete Probuilder se construyeron los cuartos de mezcla y de control, junto con algunas estructuras de ambientación. Probuilder permitió la modificación de la forma en los objetos básicos de Unity, lo que conllevó al desarrollo rápido y eficiente de prototipos. Adicionalmente, permitió el uso de herramientas exclusivas para aplicar texturas a los materiales de manera fácil. Por otra parte, además de los escenarios también se realizó el diseño de la mesa de estudio y de las bases para los monitores. La mesa se construyó con base en los modelos tradicionales de escritorio para estudios de grabación y en la siguiente imagen se observa el primer prototipo que se desarrolló para el cuarto de control junto con el diseño del escritorio. Ese prototipo inicial no aprobó las etapas de validación, fue reemplazado por otro que rescató los aciertos y mejoró los defectos.



*Figura 27 Prototipo inicial a la izquierda y a la derecha el modelo de escritorio con las bases*

*Fuente: elaboración propia*

A continuación se referencian los principales scripts que fueron creados para programar todo el sistema de mecánicas de la aplicación.

##### 4.4.1 Audio Manager

Para controlar todos los sonidos de las esferas, se diseñó un script llamado “AudioManager” el cual es el más extenso ya que controla un 50% de toda la aplicación. Es el script encargado de controlar la interfaz del Mezclador, y por tanto, gestiona funciones

vitales al momento de iniciar la aplicación. Cuando el usuario elige algunas de las opciones de Mix, el script busca los audios y los carga en un arreglo dinámico. Esa búsqueda sólo se hace cuando se accede al menú “Cargar Sonidos” y no cuando se inicia la aplicación. Por otra parte, todos los audios están ubicados en una carpeta especial de Unity llamada Resources. En Unity existen carpetas de uso especial que sirven para instanciar o acceder a los Assets de forma diferente y en diferentes tiempos. Cuando se utilizan Assets en las escenas de Unity, éstos crean instancias que luego serán compiladas al momento de exportar la aplicación. Si uno de esos Assets no pertenece a ninguna de las escenas, no se incluirá al momento de exportar. Como los audios de las esferas son Assets que no están incluidos explícitamente en un Audio Source, al momento de compilar la escena Unity no los tendría en cuenta. Por tanto los audios no podrían ser reproducidos ya que no existirían cuando se ejecute la aplicación en las gafas. Por ese motivo el uso de la carpeta Resources es fundamental, puesto que permite compilar y almacenar Assets que se pueden llamar por código en cualquier momento sin la necesidad de que se encuentren instanciados en una escena. Trabajar con Resources es de gran ayuda y se puede acceder fácilmente a los archivos tan solo con la instrucción Resources.Load. Por ejemplo, con la siguiente instrucción se cargó un clip de audio: Resources.Load<AudioClip>("subfolder/"+nombredelaudioclip), lo cual fue fácil.

#### **4.4.2 Audio Player**

Como se mencionó, la aplicación permite grabar un archivo de audio con el uso del SDK de HTC, pero por medio de los ciclos de validación y verificación se identificó la necesidad de contar con un mecanismo que logrará reproducir esas grabaciones dentro de los escenarios. Por tanto, se diseñó un script encargado de buscar los archivos y reproducirlos. Esta funcionalidad no se tenía prevista en los objetivos, sin embargo se determinó su inclusión para enriquecer las posibilidades del desarrollo. El script creado se llamó “AudioPlayer” y se encuentra en el escenario del cuarto de mezcla y el de control. Aunque la interfaz del Reproductor es bastante sencilla, el código para reproducir los archivos almacenados es bastante interesante porque se usaron varias funcionalidades de Unity. Principalmente, se utilizó otra carpeta especial llamada Streaming Assets<sup>77</sup>. Esa carpeta

---

<sup>77</sup> <https://docs.unity3d.com/es/current/Manual/StreamingAssets.html>

guarda los Assets en su formato original sin cambiarlos o comprimirlos al momento de compilar el programa cuando se está exportando. Lo que permitió tener acceso directo a archivos en su formato original (sin la posibilidad de modificarlos, solo acceso de lectura) y así mismo, permitió guardar nuevos archivos. Cuando se guardan los audios, estos se almacenan en un subfolder dentro de la carpeta de Assets Streaming. Para lograr reproducirlos de nuevo, se tuvo que utilizar al menos dos librerías básicas. Primero la librería de System.IO, la cual permitió tener acceso a ficheros del sistema operativo y la librería de Unity Engine Networking, la cual funciona para cargar archivos de un servidor en internet o de manera local por medio de instrucciones WebRequest<sup>78</sup>. La manera de acceder a los archivos es similar al Resources, la diferencia inicial radica en la forma en la que el motor compila las instancias y la sintaxis. Por ejemplo para cargar los archivos grabados se utilizó la siguiente instrucción: `Application.streamingAssetsPath + "/subfolder/"`. Utilizar las carpetas especiales de Unity expande las herramientas de gestión de ficheros para el desarrollador, por lo cual se recomienda su debida implementación.

La programación del Audio Player fue un reto e implicó bastante investigación y al menos dos prototipos fallidos. La información que se carga en la interfaz es totalmente instanciada en tiempo real mediante instrucciones de código, lo que significa que solo aparece cuando se ejecuta la aplicación. Cuando se cierra el reproductor, se destruyen los objetos y cuando se abre de nuevo, se vuelven a cargar.

#### **4.4.3 Script reproductor de video**

Para encontrar la manera más apropiada de mostrar al usuario el tutorial de manejo de la aplicación se realizaron dos prototipos. El primero utilizó una lógica poco intuitiva por lo cual se decidió que la mejor forma para lograrlo era mediante la reproducción de videos cortos que mostrarán las manos presionando los botones en los controles. Por lo cual se hizo una sesión de grabación de video para capturar el movimiento de las manos con los controles. En el material fílmico por ejemplo, se muestra que para abrir y cerrar el menú de mano se deben presionar unos botones específicos y que para agarrar objetos se debe presionar el botón de agarre y así con todas las demás funciones. Los videos fueron grabados originalmente con una cámara Nikon d5300 y la edición se realizó en el software Davinci

---

<sup>78</sup> <https://docs.unity3d.com/ScriptReference/Networking.UnityWebRequest.html>

Resolve. Para reproducir estos videos, también se utilizó la carpeta especial de Streaming Assets en donde por medio de una textura en un Raw Imagen de Unity se reproduce el video según la opción elegida.

#### **4.4.4 Script para agarrar objetos**

La aplicación cuenta con la opción para agarrar las esferas de sonido y moverlas a una posición específica. Pero también cuenta con la opción para agarrar esas esferas y acercarlas o alejarlas e incluso, rotarlas en su propio eje. Esta función también está disponible para las guías dentro del cuarto de mezcla. Para agarrar y hacer la traslación de los objetos, es necesario presionar el botón de agarre y mover el joystick en los controles. Los scripts encargados de esta mecánica se llaman “distancegrabwithoutattract” y “probe”. El primer script se añadió como componente a las esferas y a las guías, mientras que el segundo objeto se añadió a dos GameObjects encargados de controlar el sistema de entradas. El sistema de entradas constantemente revisa cuando el usuario presiona un botón y gestiona el evento que debe accionar. Cuando se mueve hacia adelante o hacia atrás la palanca de los controles, el script calcula la distancia entre el usuario y el objeto, luego cambia el Parent del objeto seleccionado hacia el Parent del control y resta o suma valores a la magnitud de la distancia calculada. Igualmente para la rotación de los objetos, una vez se mueva la palanca de los controles horizontalmente, se aplica el método Transform.Rotation.

#### **4.4.5 Otros scripts**

Los anteriores Assets y scripts fueron los más importantes para garantizar las mecánicas de trabajo en la aplicación. Sin embargo también se diseñaron scripts para el movimiento en 360 grados de las esferas, para cambiar las escenas por medio de los menús, para controlar el sonido de fondo cuando no se está en el mezclador, para controlar la simulación de sonido 3D en la parte de “Conceptos” dentro del escenario del Aula. Se diseñaron animaciones para controlar la opacidad de ciertos materiales en algunos momentos como por ejemplo en el cambio de escenas, mediante fade in y fade out de la cámara del usuario. En ese mismo escenario del Aula, el piso tiene un Shader donde aparecen cuadros con líneas que se iluminan y se oscurecen, eso se logró mediante la herramienta de Unity Shader Graph. Entre otras funciones diseñadas, lo expuesto hasta el momento da cuenta del trabajo realizado para obtener los resultados que se evidencian en la siguiente sección.

## 5 Resultados

La presente investigación resultó en un desarrollo tecnológico enfocado hacia la mezcla de sonido espacial dentro de un ambiente de Realidad Virtual. La aplicación resultante es un esfuerzo por hallar distintas interfaces de usuario entre nuevos dispositivos electrónicos y la forma tradicional de mezclar audio. En lo siguiente se presentan los escenarios y funcionalidades con las que cuenta la aplicación, en sincronía con el cumplimiento de los objetivos planteados.

### 5.1 Descripción general

3DMIXR es una aplicación en Realidad Virtual basada en el uso de representaciones visuales que permite mezclar interactivamente audio 3D. Su diseño alusivo a los estudios de producción musical presenta al usuario una experiencia inmersiva única. En donde es posible modificar la posición de diez pistas de audio individuales, dentro de un cuarto de mezcla diseñado para el libre movimiento de los sonidos. Cuenta con funciones de grabación y reproducción de las mezclas que se realizan dentro de la aplicación, junto con cinco distintas opciones de reverberación de sala, las cuales cambian la forma en la que se perciben los sonidos durante la mezcla. La aplicación contiene tres opciones de mezcla distintas y un apartado dedicado exclusivamente para mostrar el uso correcto de los controles por medio de videos cortos tipo tutorial. Además, cuenta con un espacio diseñado para percibir el sonido espacial de manera simulada mediante fuentes sonoras que aparecen aleatoriamente. En la figura 28 se observa el escritorio del estudio dentro del cuarto de control.



*Figura 28 Cuarto de control, escenario principal*

*Fuente: elaboración propia*

La parte central de la aplicación se trata de un entorno interactivo llamado “cuarto de control” que brinda al usuario una experiencia de presencialidad dentro de un estudio de grabación, el cual se encuentra ambientado con equipos de audio e instrumentos musicales. La parte principal es un entorno de mezcla de sonido espacial llamado “cuarto de mezcla”, el cual presenta varias interfaces para que el usuario pueda manipular cada sonido y también la reverberación del cuarto. Dentro del cuarto de mezcla, el usuario cuenta con la función de grabación la cual permite almacenar en un archivo de formato WAV las mezclas que se realizan. También dentro del cuarto de mezcla se encuentra un micrófono estático, con la forma de una cabeza para representar el punto de escucha sobre el cual se proyectan los sonidos. Por último, se tiene un entorno llamado “Aula” en donde se muestra el manejo de la aplicación por medio de videos que indican las acciones que se deben llevar a cabo en los controles para interactuar con los menús y con los sonidos. Allí se muestran algunos ejemplos de sonido espacial y de las herramientas de mezcla visual que tiene la aplicación. El escenario del Aula es diferente a los otros dos, pues es un ambiente que no está encerrado, es decir que no tiene límites visibles. Está pensado para involucrar al usuario en un ambiente de percepción informativa distinto, ya que es la parte de la aplicación diseñada especialmente para visualizar información útil sobre el proyecto.

Por otra parte, la aplicación se llamó 3DMIXR. Este nombre es una combinación que relaciona el sonido espacial (3D), la mezcla de sonidos (MIX) y la realidad extendida (XR). De acuerdo con lo anterior, se eligió esta denominación ya que su significado está directamente relacionado con la temática del desarrollo. En el anexo 1 se encuentra el diseño realizado para el Logotipo.

## **5.2 Escenarios e interfaz de usuario**

### **5.2.1 Cuarto de control**

Este entorno es el principal y por tanto se buscó que fuera el más llamativo. Luego de varios prototipos se llegó a determinar que la ambientación por medio de equipos de audio e instrumentos musicales era la mejor manera de transmitir al usuario la temática de la aplicación. A parte de los elementos de ambientación, también hay guías para que el usuario pueda desplazarse por los escenarios, junto con puntos de información en donde el usuario

puede indagar sobre el significado de los equipos. En la figura 29 se muestra el cuarto de control con los detalles mencionados.



*Figura 29 Cuarto de control*

*Fuente: elaboración propia*

Cabe resaltar que éste entorno tiene dimensiones hechas a escala. Es decir, el tratamiento acústico del cuarto se basó en el diseño Hipersonic Studio<sup>79</sup>, en donde se tuvieron en cuenta los desacoples de las estructuras, los materiales acústicos absorbentes, los difusores y las trampas de bajos en las esquinas como se muestra en la figura 30.

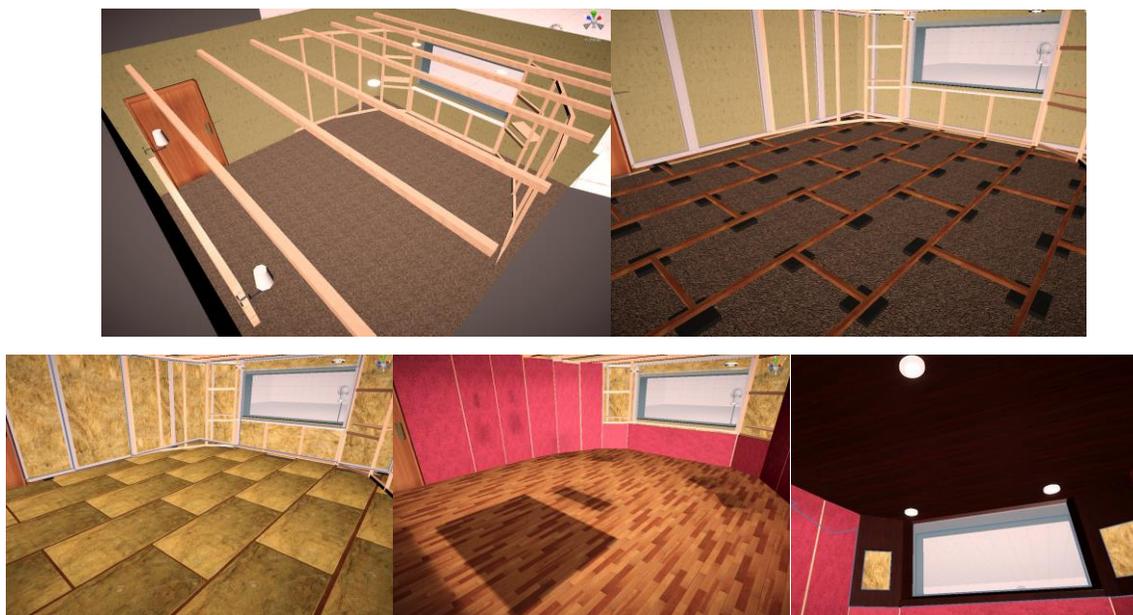


Figura 30 Simulación de acondicionamiento acústico.<sup>80</sup>

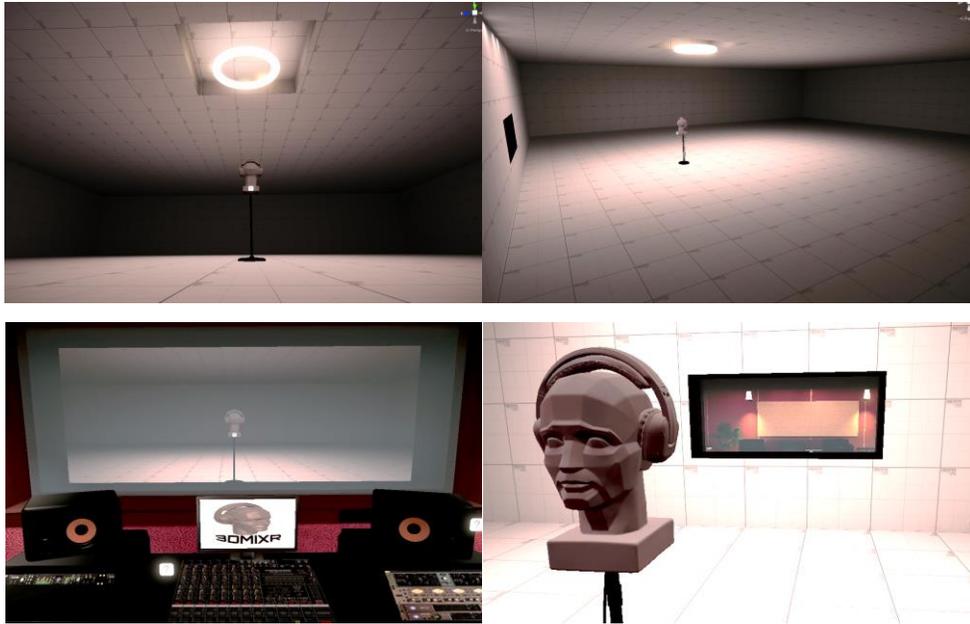
Fuente: elaboración propia

### 5.2.2 Cuarto de mezcla

A diferencia del anterior escenario este entorno tiene un diseño minimalista. Ya que se determinó que de esta forma el usuario se podría enfocar de manera más directa en la mezcla espacial de los sonidos y en su percepción. El escenario está al otro lado del cuarto de control, tras la ventana. Eso se propuso con el fin de brindar al usuario la sensación real de estar en un estudio en donde la cabina de grabación está al frente de la consola de mezcla. En este caso la cabina de grabación es un cuarto amplio en donde está ubicado un micrófono binaural el cual representa el punto de escucha con respecto a los sonidos y también, hace analogía a las cabezas de maniquí usadas en las cámaras anecoicas para obtener los datos de las HRIR.

<sup>79</sup> Ing. Ulises Trujillo Martínez. <https://www.youtube.com/watch?v=N8A7N6XIpxk>

<sup>80</sup> Este tema da para toda una investigación, que podría tratarse solamente de las posibilidades que permite la Realidad Virtual en el diseño de recintos acústicamente tratados.

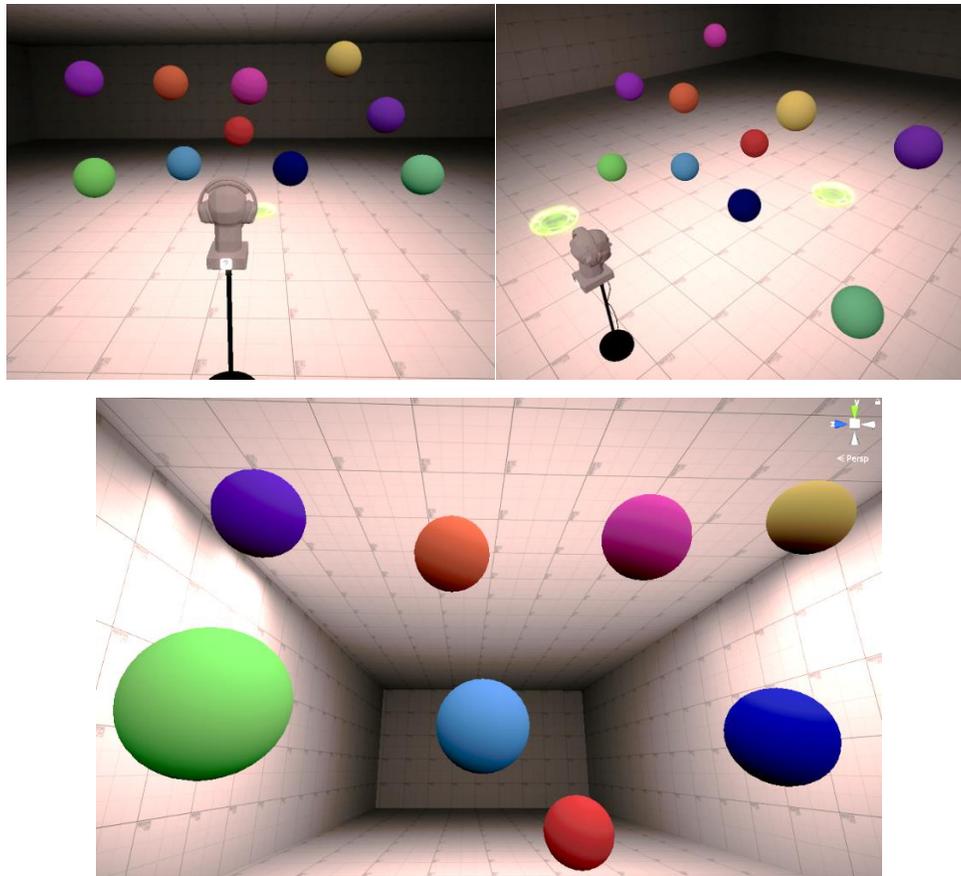


*Figura 31 Cuarto de Mezcla.*

*Fuente: elaboración propia*

### **5.2.3 Representación de los sonidos**

De acuerdo con (Agostini, 2014) los sonidos se representaron de la manera más sencilla posible para permitir al usuario un manejo simple de la interfaz. Cada sonido está representado por una esfera, la cual tiene un color distinto de acuerdo con cada canal individual en el Mezclador. El usuario puede apuntar a cada esfera y por medio de los controles es posible trasladar los sonidos a cualquier lugar dentro del cuarto de mezcla. Una esfera es un sonido espacializado, por tanto su ubicación se puede percibir respecto a la posición del punto de escucha. Con relación al punto de escucha se determinó que al tener un punto fijo de referencia para capturar los sonidos, se obtenía una grabación más precisa de la mezcla. Ya que si el punto de escucha estuviera anclado al movimiento de la cabeza del usuario, como si fueran los oídos, los sonidos podrían ser afectados por las oclusiones de los objetos en el escenario y no se obtendría una grabación consistente.



*Figura 32 Representación de los Sonidos*

*Fuente: elaboración propia*

#### **5.2.4 Aula**

Principalmente, el escenario tiene un concepto de terreno abierto (ver figura 33). Es decir, que no tiene límites en el sentido en que el usuario no se encuentra dentro de un recinto cerrado como un salón de clase por ejemplo, ni como el cuarto de control o de mezcla los cuales son cuartos con paredes y techos. El Aula tiene un diseño con una gama de colores violeta y un sistema de partículas en la parte superior, esto con el fin de transmitir información sobre la aplicación de una forma distinta a un salón convencional. Dentro del escenario se encuentra un pequeño menú de navegación que le sirve al usuario para desplazarse por las principales opciones. Aquí se presentan accesos directos a los otros entornos y a la sección de créditos, así también como una parte dedicada exclusivamente para mostrar el manejo de la aplicación por medio de videos tipo tutorial y también para mostrar una pequeña simulación relacionada con la percepción del sonido espacial. En la siguiente figura se pueden observar sus aspectos más relevantes.

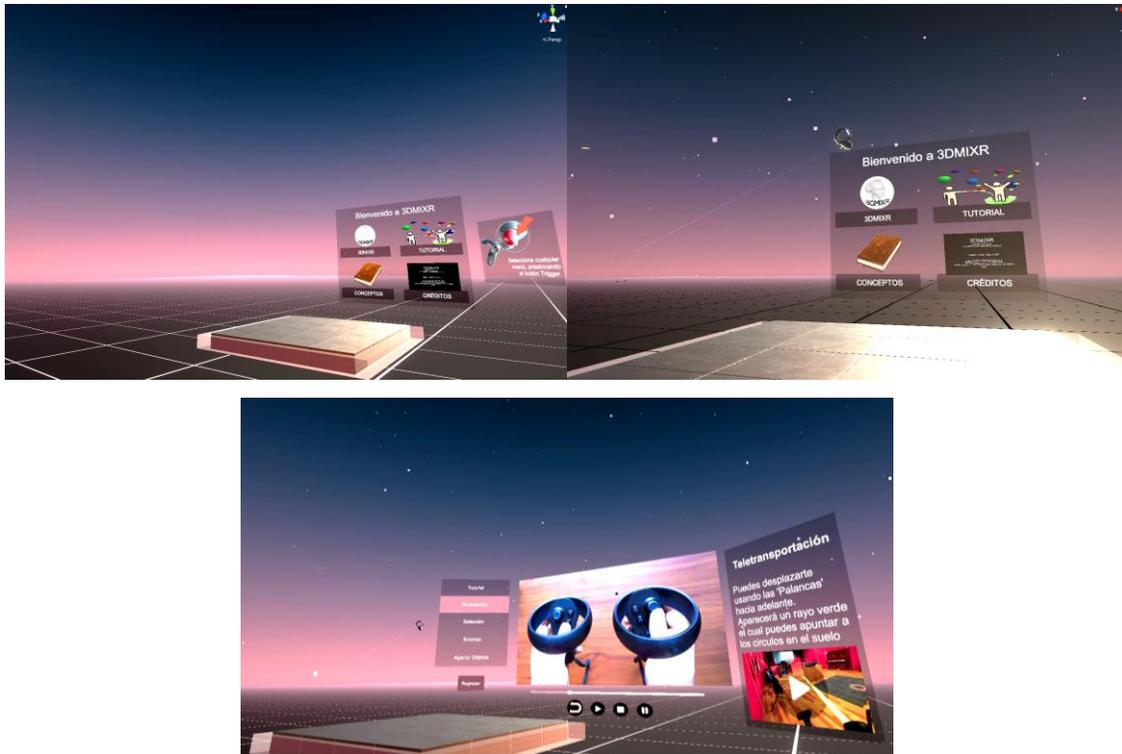


Figura 33 Entorno Aula

Fuente: elaboración propia

### 5.2.5 Menús de mano y menús móviles

La forma básica de interacción con la aplicación es mediante los menús. En el entorno de Aula hay accesos directos a los principales sitios de la aplicación, pero también es posible acceder a éstos mediante una función que se llamó el “menú de mano”. Este menú tiene una idea de fondo y es la de simular la tabla que usan los pintores en su mano para las paletas de colores. En la figura 34 se puede observar el diseño del menú y sus niveles los cuales permiten acceder a todas las funciones del escenario de mezcla y al desplazamiento entre escenas. Cuando se presiona el botón X en el control izquierdo, se abre un menú que está anclado a la mano, el cual se presenta de forma semi circular y está dividido en niveles como se puede ver en la siguiente figura. El primer nivel sirve para desplazarse por los distintos entornos. La opción “Cuarto de Control” transporta al usuario al frente del escritorio del estudio. La opción “Cuarto de Mezcla” traslada al usuario al frente del cuarto de mezcla y a su vez, activa el segundo nivel, el cual contiene opciones para mezclar los sonidos. En ese segundo nivel se incluyen opciones para cargar diferentes sonidos, para abrir el Mezclador, para reproducir

las mezclas que se hagan con la aplicación, para abrir la interfaz de transporte<sup>81</sup>, para exportar el audio en formato WAV y cambiar la reverberación del cuarto de mezcla. En un tercer nivel se encuentran las opciones para cargar sonidos y en el caso del menú de reverberación, las distintas opciones de reverb que se pueden elegir. El menú de “Aula” lleva al usuario directamente al entorno del Aula y el menú de “Cerrar App” permite al usuario salir de la aplicación.



Figura 34 Menú de mano

Figura: elaboración propia

Los menús móviles se pueden mover con el usuario sin estar anclados a la mano. Es decir, cuando los menús están activos y el usuario se desplaza por el cuarto de mezcla o el cuarto de control, los menús se mueven con respecto a la posición del usuario. Se activan por medio del menú de mano en el segundo nivel con las opciones de “Transporte” y “Exportar .WAV” (ver figura 35). La opción de exportar sirve para nombrar el archivo de audio que se desea grabar, e incluye un teclado en donde se ingresa el nombre del audio y un botón de “OK” para cerrar el menú cuando se haya terminado de escribir. Los menús móviles se pueden ubicar en cualquier parte del entorno, solo es necesario apuntar con el control hacia alguna parte del menú y con el botón de agarre, sujetarlo hasta ubicarlo en la posición deseada. Con el joystick es posible acercar y alejar los menús, así también como rotarlos sobre su propio eje. Esas opciones de movimiento aplican también para los avisos informativos que se encuentran tanto en el cuarto de control como en el cuarto de mezcla.

<sup>81</sup> La barra de transporte es un término utilizado en la producción musical para referirse a las opciones de play, stop, pause, adelante y atrás, así también como las opciones de grabación e incluso de tiempo y tiempo de la canción.

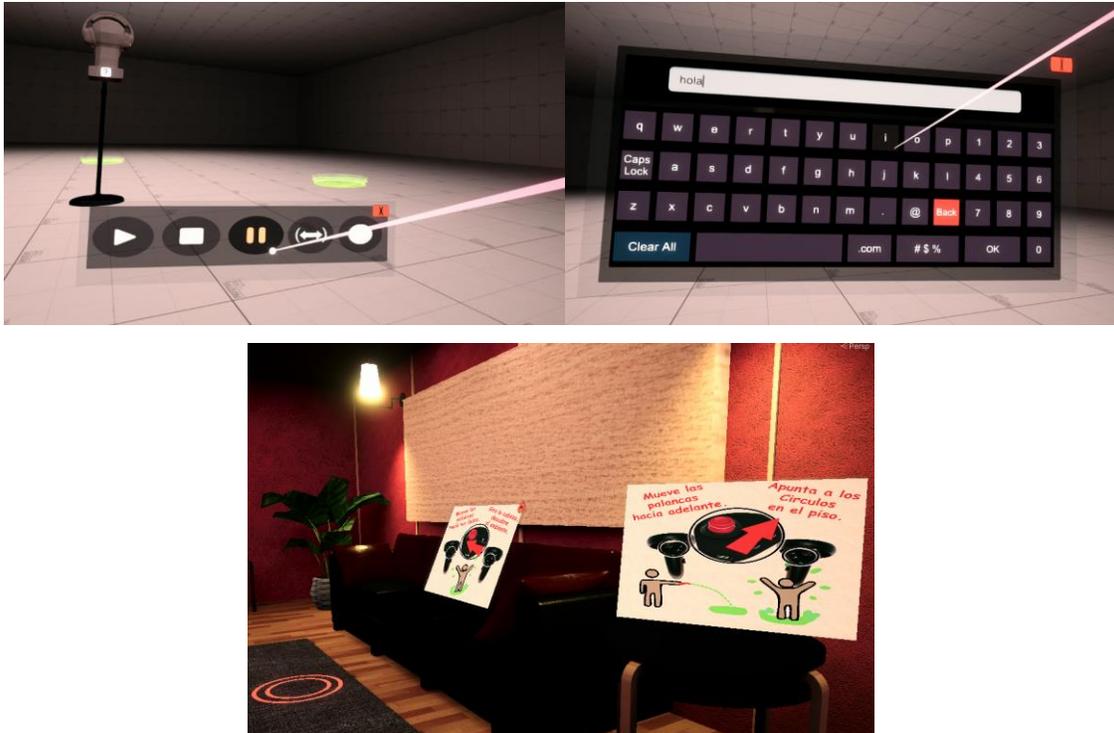


Figura 35 Menús móviles, a la izquierda la barra de transporte, a la derecha el teclado virtual y en el inferior, las guías de usuario.

Fuente: elaboración propia

### 5.3 Mecánica de la aplicación

El usuario puede interactuar con las esferas de sonidos mediante los dos controles de las gafas Oculus Quest. Dentro de la aplicación, la posición de las manos del usuario está representada por estos controles, y de ellos sale un rayo que permite apuntar y seleccionar tanto a los menús, como a las esferas de sonido y a los menús móviles. Estos últimos son susceptibles de ser ubicados donde el usuario desee. Para ello se deben pulsar los botones de agarre o el botón de gatillo en el caso de selección de menús.

#### 5.3.1 Sistema de navegación

La técnica utilizada para el desplazamiento permite al usuario moverse por los entornos virtuales apuntando a ciertas zonas en el piso. Las zonas son círculos que se iluminan al ser apuntados por el rayo que sale de los controles. Cuando el usuario apunta a estos círculos y mueve el joystick hacia adelante y lo suelta, se teletransporta a la posición donde se encuentra el círculo. En la figura 36 se muestran los dos tipos de círculos que se encuentran en la

aplicación, los cuales difieren en su color para denotar que unos están dentro del cuarto de control y los otros en el cuarto de mezcla. También se puede observar el rayo Teleport.

Por otra parte, la navegación entre Escenas se realiza mediante los menús de mano como ya se indicó antes. Por ejemplo, cuando el usuario está en el cuarto de control es posible trasladarse al cuarto de mezcla o al aula, y de igual manera desde estos sitios es posible ir a los demás.



*Figura 36 Puntos de movimiento en el Sistema de Navegación*

*Fuente: elaboración propia*

### **5.3.2 Sistema de mezcla**

El sistema de mezcla se basa en el movimiento de las esferas por medio de la selección y agarre con los controles, ya que de esta manera es posible ubicar las fuentes sonoras en una posición específica. Pero adicionalmente, la aplicación cuenta con una interfaz especial que permite acceder a otro tipo de funciones para cada sonido por separado. Esa interfaz se conoce como “Mezclador”.

#### **5.3.2.1 Mezclador**

Es una interfaz basada en la idea tradicional de las consolas de mezcla, las cuales se presentan en forma de Channel Strip. Por tanto, el Mezclador cuenta con diez canales individuales que permiten modificar el volumen de cada pista de audio, aplicar filtros frecuenciales y activar o desactivar movimientos circulares automáticos. En la siguiente figura se muestra en detalle las divisiones de la interfaz del Mezclador. Cuando se accede a

la opción “Cargar Sonidos” en el menú de mano, aparecen 3 opciones de “Mix”<sup>82</sup>. El primer Mix se refiere a una sola esfera con una pista de audio. El segundo Mix utiliza 5 canales del Mezclador y por tanto aparecen 5 esferas. Mientras que el último Mix utiliza 10 canales y aparecen 10 esferas<sup>83</sup>.

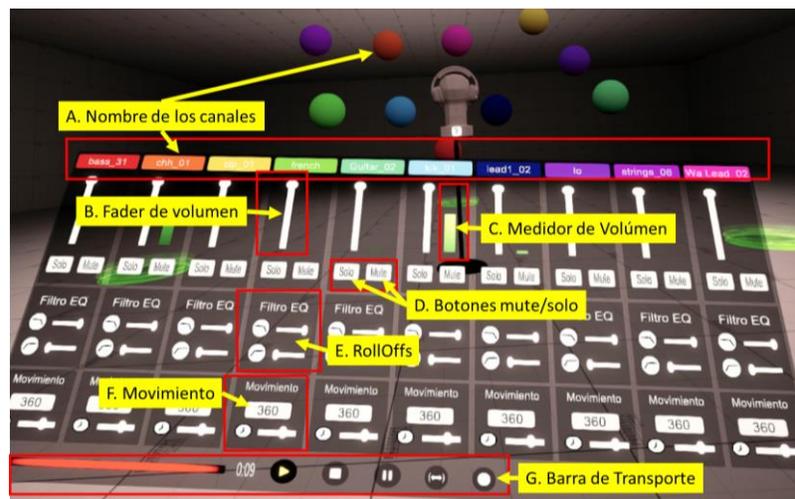


Figura 37 Interfaz del Mezclador

Fuente: elaboración propia

Una vez que el usuario ha elegido el Mix, esa información es utilizada para cargar los audios correspondientes y por tanto se modifican los nombres en la parte A del Mezclador. La distinción entre colores sirve para diferenciar un sonido de otro y son los mismos colores que tienen las esferas, por tanto cada esfera representa un canal en el Mezclador. La parte B de la figura anterior, apunta al fader de volumen de cada canal, el cual modifica el nivel de amplitud de cada audio. La parte C del Mezclador es un medidor de volumen que indica el nivel de intensidad de los sonidos. En la parte D se encuentran los botones de solo y mute, comunes en una consola de mezcla y en los DAW. Sirven para que solo suene un canal o para que suenen todos los canales menos el que este en mute. La parte E se refiere a dos filtros, uno pasa bajas y otro pasa altas que se conocen como RollOffs, los cuales cortan la parte baja o la parte alta de las frecuencias en cada canal del Mezclador. La parte F activa o desactiva un movimiento de 360 grados en las esferas alrededor del punto de escucha, y

<sup>82</sup> Como los requisitos de la aplicación estaban acotados a un máximo de 10 sonidos simultáneos, se determinó que tener 3 opciones de mezcla sería lo mejor para probar el funcionamiento del Mezclador.

<sup>83</sup> La música utilizada para los Mix no tiene restricciones de copyright, por lo cual es posible su uso comercial o académico.

cuenta con un deslizador que permite variar la velocidad de rotación. Por último, en la parte G se encuentra una barra de transporte similar a la del menú móvil. La diferencia es que al lado izquierdo hay una barra roja que indica el progreso y duración de los sonidos en el mezclador.

### 5.3.3 Grabación y reproducción

Una de las funciones que más se destacan de la aplicación es la posibilidad de grabar y reproducir las mezclas que el usuario realiza dentro del cuarto de mezcla. Para ello, se puede acceder al menú de mano para abrir el teclado virtual, por medio del cual se nombra el archivo que se desea guardar. Luego, en cualquiera de las dos barras de transporte se activa el botón de “Rec”, el cual inicia la grabación hasta que el usuario detenga el proceso de almacenamiento. Una vez guardado el archivo en formato WAV, es posible abrirlo mediante el acceso a la opción “Reproductor” en el menú de mano. El Reproductor permite encontrar las grabaciones hechas con la aplicación y las muestra al usuario en forma de lista. En la siguiente figura se muestra la interfaz del Reproductor en donde se muestran los botones básicos para reproducir las mezclas.

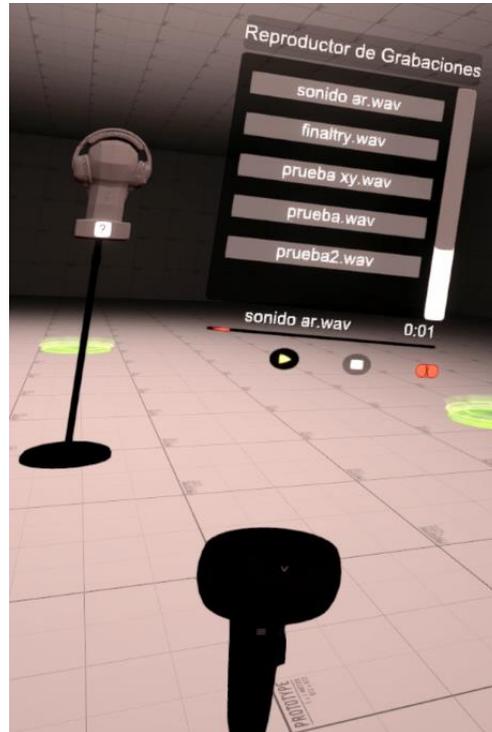


Figura 38 Menú Reproductor de grabaciones

Fuente: elaboración propia

### 5.3.4 Reproducción de video clips

El sistema de reproducción de videos se encuentra dentro de las opciones de “tutorial” en el entorno del Aula. Es una interfaz que permite reproducir videos cortos que muestran los botones que se deben pulsar para realizar una determinada acción dentro de los escenarios de cuarto de control y cuarto de mezcla. El diseño de este apartado se enfoca en ayudar al usuario a familiarizarse con las opciones básicas de desplazamiento e interacción que ofrece 3DMIXR.



Figura 39 Reproducción de Videos

Fuente: elaboración propia

### 5.3.5 Simulación del concepto de audio 3D

En la parte de “Conceptos” en el escenario del Aula se encuentra una animación diseñada para explicar la idea de sonido espacial mediante la aparición aleatoria de fuentes sonoras en distintos tiempos. El objetivo es que el usuario pueda descubrir la ubicación de los sonidos al momento en que estos aparecen. La animación comienza con una voz femenina de un avatar en frente del usuario. Luego continúa hablando otro avatar al lado izquierdo, seguido de otro en la parte de atrás y otro a la derecha. Mientras hablan van explicando el concepto de sonido espacial y al final de la información se invita al usuario a detectar la posición de las fuentes, mientras se reproduce el sonido del motor de un carro, el disparo de un cañón de un carro tanque, el gruñido de un león y un sonido Ambisonic.



*Figura 40 Simulación de sonido espacial*

*Fuente: elaboración propia*

### **5.3.6 Traslación y Rotación de los Objetos**

Cuando el rayo que sale de los controles apunta a una esfera o uno de los menús móviles, es posible mover estos objetos presionando el botón de agarrar. De esta manera se pueden mover las esferas y los menús de acuerdo con la posición de las manos, por ejemplo. Pero otra función interesante es la de agarrar los objetos y por medio del joystick acercarlos, alejarlos o rotarlos. Esto se consigue presionando el botón de agarrar simultáneamente con los joystick. Si se mueve la palanca hacia adelante el objeto se va a alejar, y si la palanca se mueve hacia atrás, se acerca el objeto hacia el usuario. También se puede rotar el objeto si se mueve la palanca hacia los costados izquierdo o derecho.

## 6 Conclusiones

Con base en lo expuesto en la sección de resultados, es posible afirmar que el cumplimiento total de los objetivos propuestos permitió la obtención de un desarrollo tecnológico con un nivel superior a lo esperado. Puesto que se incluyeron funciones y mecánicas que no estaban contempladas en los requisitos, pero que aportaron de manera significativa en el resultado final.

Los resultados obtenidos en este desarrollo permitieron vislumbrar el potencial que ofrece la Realidad Virtual en relación con el procesamiento de audio 3D. Puesto que la espacialidad de los sonidos se puede percibir con gran agudeza si se cuenta con un dispositivo que permita Head Tracking, como en las gafas Oculus Quest. Por otra parte, el alcance de esta investigación también mostró la posibilidad de obtener nuevas interfaces de usuario que logren cambiar la forma tradicional de mezclar canciones basadas en la idea de Channel Strip de las consolas de mezcla convencionales.

El desarrollo de los escenarios virtuales resultó en la creación de tres entornos inmersivos, los cuales permitieron estudiar a fondo las ventajas que ofrece la Realidad Virtual aplicada a temas relacionados con producción musical, en especial con Audio 3D. Por tanto, el diseño del cuarto de control fue pensado para brindar al usuario una experiencia en la cual se pudiera estar dentro de un estudio de grabación, al lado de los equipos, instrumentos y consolas que usualmente se utilizan en estos espacios. Mientras que el escenario del cuarto de mezcla se pensó para que no existieran distracciones y se le permitiera al usuario un enfoque hacia la mezcla de los sonidos.

Por otra parte, las dimensiones del cuarto de mezcla al ser diseñadas para que las esferas de los sonidos tuvieran un mayor grado de libertad en sus movimientos permitió identificar la capacidad que tiene la aplicación para motivar el arte al momento de mezclar interactivamente los sonidos individuales. Y en cuanto al diseño del escenario del aula, es posible argumentar que se tuvieron en cuenta aspectos estéticos que permitieron transportar al usuario a un lugar distinto a los otros escenarios. Puesto que la intención principal fue comunicar de manera distinta la información sobre las funciones de los controles y los conceptos de audio espacial que utiliza la aplicación.

Sobre las mecánicas de interacción del usuario con respecto a los escenarios e interfaces gráficas, se logró constatar que la técnica Teleport reduce significativamente los efectos de Motion Sickness, permitiendo un desplazamiento del usuario sin implicaciones negativas en su salud. Por otra parte, el diseño de los menús fue creado para organizar las distintas opciones de la aplicación de forma simple pero eficaz, por tanto el prototipo seleccionado fue el menú de mano, ya que permitió un acceso rápido a todas las funciones del escenario de mezcla y también a las opciones de navegación entre entornos. El menú de mano fue un elemento sumamente importante dentro de toda la mecánica de la aplicación, puesto que es el encargado de generar el acceso al Mezclador y a las demás funcionalidades, allí es donde se encuentran los parámetros de control y navegación para todos los escenarios.

El sistema de grabación de las mezclas significó un avance considerable en el uso de la Realidad Virtual orientada hacia la producción musical, ya que una de las funciones básicas de todo DAW es la renderización de los proyectos musicales. Por otra parte, la grabación de las mezclas implicó el desarrollo de un reproductor, por lo cual se tuvo que modificar los requisitos del proyecto para que también se contemplara la creación de un sistema nuevo, que permitiera acceder a los ficheros para que el usuario lograra su reproducción. La obtención de este sistema también significó un gran avance para pensar en un DAW en RV.

En un sentido estricto, la aplicación en Realidad Virtual que se logró por medio de esta investigación permitió ampliar la comprensión sobre la grabación de mezclas de pistas de audio individuales dentro de un ambiente virtual que contó con las opciones necesarias para modificar la ubicación espacial de los sonidos. Y aunque ya se habían hecho aplicaciones bajo el concepto de mezcla visual, este desarrollo propuso una nueva manera de trabajar la producción musical desde la Realidad Virtual, por medio de nuevas interfaces gráficas de usuario. Por otra parte, este trabajo se trata de la primer investigación de la Universidad Pedagógica Nacional realizada sobre audio 3D o sonido espacial aplicado en una de las plataformas para Realidad Extendida. Por tanto, es un avance significativo para los semilleros de investigación de la Licenciatura en Electrónica, para el departamento de tecnología y la universidad.

Finalmente, de esta investigación fue posible vislumbrar una serie de recomendaciones y trabajos futuros que pueden aportar nuevos conocimientos al producto desarrollado, así

también como la generación de nuevas investigaciones que contribuyan al avance académico personal, de la Universidad y el país. En ese sentido, la aplicación desarrollada buscó un alto grado de realismo en los gráficos y de precisión en cuanto a la espacialización de los sonidos, por tanto el consumo de recursos no fue compatible para dispositivos móviles ya que en este caso, fue la computadora la que se encargó del procesamiento. Sin embargo, las gafas Oculus Quest permiten cargar aplicaciones en su sistema operativo sin la necesidad de una computadora. Por lo tanto, un trabajo futuro podría ser la optimización de la aplicación hacia dispositivos móviles. Lo cual implica cambiar el nivel de realismo de los gráficos y la búsqueda del menor consumo computacional por parte de los algoritmos de espacialización del audio 3D. Para ello, el SDK de Resonance audio podría ser una opción si se solucionasen los limitantes del Audio Mixer y también si se logrará un acceso a la carpeta de Streaming Assets sin hacer uso de un servidor Web. Pero además de la optimización a dispositivos móviles, la aplicación tiene bastante que ofrecer para ser implementada con equipos de escritorio y laptops sin necesidad de usar gafas de Realidad Virtual. Por tanto otro trabajo de mejora puede estar enfocado en hacer que la aplicación funcione simultáneamente para RV y para PC. Por otra parte, una trabajo posterior sobre la imagen de los sonidos es algo que puede enriquecer de sobre manera el prototipo realizado. Lo que implicaría toda una investigación del tema, puesto que la imagen de un sonido tiene que ver con la subjetividad psicológica del oyente, la memoria y las experiencias, las características físicas de los sonidos, la respuesta frecuencial de las formas anatómicas de las personas, la respuesta al impulso del espacio exterior, entre otros aspectos que generan dificultades al momento de plantear una representación visual general. Sin embargo, no es una tarea imposible y la Realidad Virtual es la mejor tecnología disponible para llevarlo a cabo. Para concluir, se recomienda a los estudiantes de la Licenciatura en Electrónica, y al Departamento de Tecnología en general, el análisis del presente trabajo, ya que se encontraron al menos dos ideas para construir trabajos de grado que pueden ser de gran valor para el grupo de investigación Alice, para la Universidad Pedagógica Nacional y en general para toda persona o institución a nivel internacional que se encuentre interesada en el audio 3D y la Realidad Virtual.

## 7 Bibliografía

- Agostini, G. (2014). *3D visual mixing Prototyping and implementation of a visual mixing tool based on natural user interfaces realtime interaction*. Roma: University of Rome II.
- Barrera Giraldo, V. (2009). *Sistema espacializador de sonido en tres dimensiones basado en la función de transferencia relacionada con la cabeza (HRTF)*. Bogotá: Universidad San Buenaventura.
- Begault, D. (1994). *3-D Sound for virtual reality and multimedia*. United States: Academic Press Inc.
- Berenguer Alarcón, J., i Castellanes, M., & Jordà Puig, S. (2005). *Música y sistemas acústicos*. Catalunya: Universitat Oberta de Catalunya.
- Bes, M. (Agosto de 2020). *Anymotion*. Obtenido de <https://www.anymotion.nl/portfolio/vr-sound-engineer-retro-recordings/>
- Brinkman, W.-P., Hoekstra, A. R., & Egmond, R. v. (2015). The effect of 3D audio and other audio techniques on virtual reality experience. *Studies in health technology and informatics*, 219, 44-48.
- Cardona Rojas, D., & Valderrama Pulgarín, M. (2014). *Desarrollo de Plugin VST, para la reproducción binaural de fuentes virtuales en movimiento, utilizando bases de datos de Funciones de Transferencia Asociadas a la Cabeza*. Medellín: Universidad San Buenaventura.
- Carrascal, J., & Jordà, S. (2011). Multitouch Interface for Audio Mixing. *Proceedings of the International Conference on New Interfaces for Musical Expression*.
- Carrión Isbert, A. (1998). *Diseño acústico de espacios arquitectónicos*. Catalunya: Universitat Politècnica de Catalunya.
- Cengarle, G. (2012). *3D audio technologies: applications to sound capture, post-production and listener perception*. Tesis doctoral Universitat Pompeu Fabra.

- Deitel, H., & Deitel, P. (1998). *Cómo programar en JAVA primera edición*. Naucalpan de Juárez: Prentice-Hall Hispanoamerica S.A.
- Gaitán Quiasua, J. (2015). *Implementación del Leap Motion como controlador midi interactivo mediante reconocimiento gestual (dactilar)*. Bogotá: Universidad San Buenaventura.
- Garcia Ruiz, J. M. (2020). *Otorrinoweb*. Obtenido de [https://www.otorrinoweb.com/images/Descargas/04.1/04.1 fisiologia\\_sistema\\_auditivo.pdf](https://www.otorrinoweb.com/images/Descargas/04.1/04.1 fisiologia_sistema_auditivo.pdf)
- Gelineck, S., & Overholt, D. (2016). Haptic and Visual feedback in 3D Audio Mixing Interfaces. *Aalborg University Copenhagen*.
- Gibson, D. (1997). *The Art Of Mixing A Visual Guide To Recording Engineering And Production*. . New York: Routledge.
- Gibson, D. (2002). *United States Patente n° US 6,490,359 B1*.
- Gupta, R., Lam, B., Hong, J.-Y., Ong, Z.-T., Gan, W.-S., Chong, S., & Feng, J. (2017). 3d audio ar/vr capture and reproduction setup for auralization of soundscapes. *24th International congress on sound and vibration*.
- Jarilla Romo, V. (2018). *Guía de desarrollo de un videojuego con Unity en realidad virtual*. Valencia: Universidad Politécnica de Valencia.
- Lanier, J. (1988). A Vintage Virtual Reality Interview. (A. Heilbrun, Entrevistador)
- Lanier, J. (2017). *Dawn of the New Everything: A Journey Through Virtual Reality*. Random House.
- Lidon, M. (2019). *Unity 3D*. México: Alfaomega.
- McCarthy, B. (2007). *Sound Systems: Design and Optimization - Modern techniques and tools for sound system design and alignment*. Focal Press.
- Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1994). Augmented Reality: A class of displays on the reality-virtuality continuum. *Telemanipulator and Telepresence Technologies. SPIE Vol. 2351*, 282-292.

- Miyara, F. (1999). *Introducción a la Psicoacústica*. ResearchGate paper.
- Oppenheim, A., & Willsky, A. (1998). *Señales y Sistemas, 2da Ed.* Prentice-Hall, Inc.
- Owens, J. (2011). *United States Patente n° US 2011/0271186A1*.
- Pérez Martínez, F. (2001). Presente y Futuro de la Tecnología de la Realidad Virtual. *Creatividad y Sociedad*.
- Pérez, B. (2015). *Introducción a la Realidad Virtual*. Universitat Politècnica de Catalunya.
- Pfleeger, S., & Atlee, J. (1998). *Software engineering: theory and practice*. India: Pearson Education.
- Prandoni, F. (2017). *A virtual reality environment with personalized spatial audio rendering*. Università degli Studi di Padova.
- Pratt, C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology (13)*, 278-85.
- Quevedo Rodríguez, J., & Malagón Barrera, B. (2019). *Desarrollo de una herramienta de apoyo para la capacitación en el manejo de extintores usando virtualidad aumentada*. Bogotá: Universidad Pedagógica Nacional de Colombia.
- Rubio Tamayo, J. (2019). Realidad extendida, interactividad y entornos inmersivos 3d: Revisión de la literatura y proyecciones. *Actas Icono 14, 1(1)*, 396-415.
- Ruiz-Velasco Sánchez, E. (2007). *Educatrónica: innovación en el aprendizaje de las ciencias y la tecnología*. Mexico: UNAM.
- Serway, R., Vuille, C., & Faughn, J. (2009). *Fundamentos de Física 8va Edición*. Cengage Learning S.A.
- Sora, C. (2017). *Una inmersión en el audiovisual VR y 360*. España: Ministerio de Economía y Competitividad.
- Soto, E., Vega, R., Chávez, H., & Ortega, A. (2003). Fisiología de la audición: la cóclea. *Instituto de Fisiología Universidad Autónoma de Puebla*.

- Systematic Sound*. (07 de Diciembre de 2020). Obtenido de <https://systematic-sound.com/3d-audio-for-sound-designers-part-i-spatial-hearing>
- Tamayo y Tamayo, M. (2003). *El Proceso de la Investigación Científica incluye evaluación y administración de proyectos de investigación científica*. Distrito Federal: Lmusa S.A.
- Vallejo Fernández, D., González Morcillo, C., Villa Alises, D., Jurado Monroy, F., Moya Fernández, F., Alonso Albusac Jiménez, J., . . . González Sánchez, J. (2014). *Desarrollo de Videojuegos: Un Enfoque Práctico*. Ciudad Real: EdLibrix.
- Vázquez Alonzo, Á., & Alarcón Zamora, M. (2010). *Didáctica de la tecnología*. Madrid: Síntesis.
- Vera Ocete, G., Ortega Carrillo, J., & Burgos González, M. (2003). La realidad virtual y sus posibilidades didácticas. *ÉticaNet*, ISSN: 1695-324X.
- Weekhout, H. (2019). *Music production: learn how to record, mix, and master music*. New York: Routledge.
- Wittek, H. (2016). ORTF-3D: Ambience recording for 3D Audio & VR. *Schoeps Mikrofone*.
- WOODCOCK, J., DAVIES, W., & COX, T. (2019). Influence of Visual Stimuli on Perceptual Attributes of Spatial Audio. *J. Audio Eng. Soc.* (67), 557–567. doi.org/10.17743/jaes.2019.0019.
- Zhang, W., Samarasinghe, P., Chen, H., & Abhayapala, T. (2017). Surround by Sound: A Review of Spatial Audio Recording and Reproduction. *Applied Sciences*.
- Zhong, X. (2015). *Dynamic Spatial Hearing by Human and Robot Listeners*. Arizona: ARIZONA STATE UNIVERSITY.

## 8 Anexos

### 8.1 Anexo 1 Logotipo



*Figura 41 Logotipo*

*Fuente: elaboración propia*

## 8.2 Anexo 2: guías de usuario parte 1.



Figura 42 tablón con las guías de usuario parte 1

Fuente: elaboración propia

8.3 Anexo 3: guías de usuario parte 2.

**¡BIENVENIDO A TÚ AMBIENTE DE MEZCLA!**

Selecciona el botón OK para continuar...

Observa la estructura del mezclador. Aquí puedes cambiar el volúmen de los canales, ajustar filtros paso alto y paso bajo. También puedes dar movimiento 360°.

En la parte de abajo del mezclador, estan los botones de play, stop, pause, loop y de rec. También conocidos como la barra de transporte. Selecciona play para comenzar!

En el menú de mano selecciona "transporte" para desvincular la barra del mezclador.

En el menú de mano selecciona "Exportar .wav". En el teclado ingresa el nombre del archivo a guardar.

En la barra de transporte selecciona el botón REC para iniciar la grabación de la mezcla que hayas elegido. Puedes mover los sonidos y modificar el mezclador, todo quedará guardado.

Selecciona el menú "Reverb Room" para cambiar la reverberación del cuarto de mezcla. Tienes 5 opciones, pruebalas.

Selecciona en el menú de mano, la opción "Aula" para cambiar al escenario en donde podrás conocer más acerca de la app.

Selecciona la opción "Cerrar App" Para salir de la aplicación.

Figura 43 tablón con las guías de usuario parte 2

Fuente: elaboración propia

#### 8.4 Anexo 4: listado de Assets y paquetes

Número	Nombre	Asset o Paquete	Categoría	Comentarios
1	VR Beginner: The Escape Room	Paquete	Gratuito	Es un proyecto de Unity Learn para aprender a usar el paquete de XR
2	ProBuilder	Paquete	Gratuito	
3	ProGrids	Paquete	Gratuito	
4	Oculus XR	Paquete	Gratuito	
5	Unity Recorder	Paquete	Gratuito	
6	Universal RP	Paquete	Gratuito	
7	XR Interaction Subsystems	Paquete	Gratuito	
8	XR Interaction Toolkit	Paquete	Gratuito	
9	XR Plugin Management	Paquete	Gratuito	
10	Resonance Audio	Paquete	Gratuito	SDK para audio 3d de Google
11	HTC Unity Plugin	Paquete	Gratuito	SDK para audio 3d de HTC
12	Bakery	Paquete	Pago	
13	UI Kit	Asset	Pago	
14	Original Wood Texture	Asset	Gratuito	
15	Acoustic Guitar	Asset	Gratuito	
16	ArchViz Sofa Pack - Lite	Asset	Gratuito	
17	Bass Guitar	Asset	Gratuito	
18	Brick Material Red Rough-Hewn	Asset	Gratuito	
19	Customizable skybox	Asset	Gratuito	

20	Customizable Stylized Brick Wall	Asset	Gratisito	
21	Door Free Pack Aferar	Asset	Gratisito	
22	Electric Guitar	Asset	Gratisito	
23	FREE Rockstar Asset Pack	Asset	Gratisito	
24	Free Rug Pack	Asset	Gratisito	
25	Golden Tiger	Asset	Gratisito	
26	Gridbox Prototype Materials	Asset	Gratisito	
27	Guitar And Amp	Asset	Gratisito	
28	HDRP Furniture Pack	Asset	Gratisito	
29	High Quality Realistic Wood Textures - Mega Pack	Asset	Gratisito	
30	HQ Acoustic system	Asset	Gratisito	
31	HQ Racing Car Model No.1203	Asset	Gratisito	
32	Keyboard Adjustable Colors	Asset	Gratisito	
33	Lamps Pack	Asset	Gratisito	
34	Microphone Mic DJ PBR	Asset	Gratisito	
35	Pack Gesta Furniture #1	Asset	Gratisito	
36	T95 Super Heavy Tank	Asset	Gratisito	
37	UMA 2 - Unity Multipurpose Avatar	Asset	Gratisito	

38	Wood Pattern Material	Asset	Gratuito	
39	Cassette Deck (Yamaha KX-202)	Modelo 3D	Gratuito	<a href="https://sketchfab.com/3d-models/cassette-deck-yamaha-kx-202-559b55501e0346eb94c5dff7452af028">https://sketchfab.com/3d-models/cassette-deck-yamaha-kx-202-559b55501e0346eb94c5dff7452af028</a>
40	AKAI MPK Mini	Modelo 3D	Gratuito	<a href="https://sketchfab.com/3d-models/akai-mpk-mini-f7a837c9b469472084940ab3d6609eb2">https://sketchfab.com/3d-models/akai-mpk-mini-f7a837c9b469472084940ab3d6609eb2</a>
41	KRK Classic 5 Studio Monitor Speaker	Modelo 3D	Gratuito	<a href="https://sketchfab.com/3d-models/krk-classic-5-studio-monitor-speaker-5c03f6b1082a4decb33ae7511209271a">https://sketchfab.com/3d-models/krk-classic-5-studio-monitor-speaker-5c03f6b1082a4decb33ae7511209271a</a>
42	Minimoog Model D Synthesizer	Modelo 3D	Gratuito	<a href="https://sketchfab.com/3d-models/minimoog-model-d-synthesizer-6d1aa9e5026d4fa68938254d082c8693">https://sketchfab.com/3d-models/minimoog-model-d-synthesizer-6d1aa9e5026d4fa68938254d082c8693</a>
43	Yamaha Dx7	Modelo 3D	Gratuito	<a href="https://sketchfab.com/3d-models/yamaha-dx7-3f80009f37d644639b4a0459c86277df">https://sketchfab.com/3d-models/yamaha-dx7-3f80009f37d644639b4a0459c86277df</a>
44	Planar head by Oleg Toropygin	Modelo 3D	Gratuito	<a href="https://sketchfab.com/3d-models/planar-head-by-oleg-toropygin-7d831ee5b5494ab595c443088e549c59">https://sketchfab.com/3d-models/planar-head-by-oleg-toropygin-7d831ee5b5494ab595c443088e549c59</a>
45	Mixer DYNACORD CMS600	Modelo 3D	Gratuito	<a href="https://sketchfab.com/3d-models/mixer-dynacord-cms600-4f08d120a3aa49e1814f0b543af23b29">https://sketchfab.com/3d-models/mixer-dynacord-cms600-4f08d120a3aa49e1814f0b543af23b29</a>
46	Keyboard (Casio VL-Tone)	Modelo 3D	Gratuito	<a href="https://sketchfab.com/3d-models/keyboard-casio-vl-tone-be22fbe1ef0044f3bedc6e3edf459593">https://sketchfab.com/3d-models/keyboard-casio-vl-tone-be22fbe1ef0044f3bedc6e3edf459593</a>

*Tabla 1 Listado de Assets y Paquetes*

*Fuente: Elaboración propia*