

METODOLOGÍA DE GRABACIÓN AMBISONIC EN CONJUNTOS MUSICALES ACÚSTICOS

Javier Rubio Arenas

Tutor: Jose Javier López Monfort

Trabajo Fin de Grado presentado en la Escuela Técnica Superior de Ingenieros de Telecomunicación de la Universitat Politècnica de València, para la obtención del Título de Graduado en Ingeniería de Tecnologías y Servicios de Telecomunicación

Valencia, 10 de noviembre de 2020 Curso 2019-20

Resumen

Este estudio final de grado consiste en la implementación de la tecnología Ambisonics con el fin de reproducir audio envolvente tridimensional. Con el objetivo de conseguir una perfecta experiencia acústica tridimensional, se graban conjuntos musicales acústicos, se convierten a formato tecnológico HRTF y se reproducen en audición binaural.

El sistema Ambisonic es una tecnología que permite la grabación, configuración y reproducción de señales acústicas en sistemas conformados por 'n' altavoces prácticos. De esta manera los receptores pueden localizar los sonidos que se ubican en el espacio de tres dimensiones.

Se estudia la metodología de grabación de grupos musicales acústicos. Primero, experimentando en los laboratorios de acústica de la universidad y posteriormente grabando escenas reales como solistas con sus instrumentos, grupos musicales al completo, bandas de música etc. Finalmente, se analiza los resultados del audio utilizando los diferentes formatos existentes (Formato-A, Formato-B, Ambix y FuMa).

En este proyecto se ha profundizado en el estudio tanto matemático como instrumental de las técnicas de grabación Ambisonics. Obteniendo así una serie de pruebas y grabaciones, que junto con el "know how", serán útiles para futuros estudios dentro del sector.

Resum

Aquest treball de final de grau consisteix en la implementació de la tecnologia Ambisonics com a solució d'àudio envoltant tridimensional. Amb l'objectiu d'aconseguir una perfecta experiència acústica tridimensional, es graven conjunts musicals acústics, es converteixen a format tecnològic HRTF i es reproduïen en audició binaural.

El sistema Ambisonic de so envoltant, és una solució tecnològica que permet gravar, manipular i àudio en sistemes de m altaveus pràctics. D'aquesta manera els oients podran percebre els sons situats en l'espai tridimensional.

S'estudia la metodologia d'enregistrament de grups musicals acústics. Primer, experimentant en els laboratoris d'acústica de l'universitat i posteriorment gravant escenes reals com a solistes amb els seus instruments, grups musicals al complet, bandes de música etc. Finalment, s'analitza la seva reproducció utilitzant els diferents formats existents (Format-A, Format-B, Fuma i ambix).

En aquest projecte s'ha profunditzat en l'estudi tant matemàtic com instrumental de les tècniques d'enregistrament Ambisonics. Obtenint així una sèrie de proves i gravacions, que juntament amb el "know how", sens dubte útils per a futurs estudis dins de el sector.

Abstract

This final degree thesis study consists of the implementation of Ambisonics technology in order to reproduce three-dimensional surround sound. With the aim of achieving a perfect three-dimensional acoustic experience, acoustic music groups are recorded, converted to HRTF technology format and reproduced in binaural hearing.

The Ambisonic system is a technology that allows the recording, configuration and reproduction of acoustic signals in systems made up of 'n' practical speakers. In this way the receivers can locate sounds that are located in the three-dimensional space.

The methodology for recording acoustic music groups is studied. First, by experimenting in the university's acoustics laboratories and then recording real scenes as soloists with their instruments, complete bands, etc. Finally, the results of the audio are analysed using the different existing formats (Format-A, Format-B, Ambix and FuMa).

In this project, both the mathematical and instrumental study of Ambisonics recording techniques has been deepened. Thus obtaining a series of tests and recordings, which together with the "know how", will be useful for future studies within the sector.

Palabras Clave

- **Ambisonics**
- **Binaural**
- **Reaper**
- **Plugins**
- **Renderizar**
- **Filtros**
- **Rode**

Índice

Capítulo 1. INTRODUCCIÓN AL FORMATO DE SONIDO ENVOLVENTE AMBISONICS	3
1.1 Métodos de grabación	4
1.1.1 Introducción a la metodología de grabación	4
1.1.2 Tipos de Microfonía y técnicas de grabación.....	6
1.1.3 Configuración de Formatos. A y B	9
1.2 F. Codificación	13
1.2.1 Señales Ambisonics, Estéreo y Formato UHJ.....	13
1.3 F. Decodificación	14
1.4 Física de la audición binaural.....	15
1.4.1 Psicoacústica de la localización espacial.....	16
1.4.2 Indicadores de elevación y HRTF	18
1.4.3 Indicadores de distancia y vector distancia.	19
1.5 Respuestas binaurales.....	19
1.6 Aplicaciones	20
1.7 Conversión a binaural.....	20
Capítulo 2. Hardware	21
2.1 Dispositivos.....	21
2.1.1 Micrófono Rode NT-SF1	21
2.1.2 Tarjeta de Audio Roland Octa-Capture. Modelo UA 1010.....	24
Capítulo 3. Software empleado en las medidas	25
3.1 Aplicaciones.....	25
3.1.1 Software de procesado. Reaper Digital Audio.	25
3.1.2 Plugins ATK.....	25
3.1.3 RØDE PLUGIN B to Ambix	26
Capítulo 4. Metodología y Pruebas realizadas.....	29
4.1 Pruebas realizadas en laboratorio	29
4.1.1 Procedimiento Experimental en laboratorio.....	29
4.1.2 Grabaciones experimentales fuera de laboratorio	39
4.2 Grabaciones a conjuntos acústicos musicales.	42
4.2.1 Laura Moreno	43
4.2.2 Julián Redondo.....	46
Capítulo 5. Conclusiones	53



Capítulo 6.	BibliografíaC.....	54
Capítulo 7.	Anexos.....	55
7.1	Anexo A: Archivos Reaper en formato .rpp.....	55
7.2	Anexo B: Canciones Binaural.....	55
7.3	Anexo C: Fotos y Vídeos	55

Capítulo 1. INTRODUCCIÓN AL FORMATO DE SONIDO ENVOLVENTE AMBISONICS

La tecnología Ambisonics trata de un formato de sonido envolvente de esfera completa, cubriendo todas las fuentes de sonido por encima y por debajo del oyente además del plano horizontal. Esta solución capacita al usuario para trabajar en grabación, manipulación y reproducción de audio en sistemas de altavoces convencionales para que los receptores del sonido puedan ubicar estos en el espacio tridimensional.

[1] El sistema de sonido Ambisonics ofrece una solución tecnológica al problema de codificar direcciones y amplitudes de la señal acústica y reproducirlas, simulando así la referencia original de las fuentes. Esto ocurre a través de un escenario horizontal de 360 grados (sistemas pantofónicos) o a lo largo de una esfera completa (sistemas perifónicos).

[2] En los años 70 un equipo de investigadores de nacionalidad inglesa, entre los que se encontraba Michael Gerzon, Peter Barnes Fallgett y John Stuart Wright, diseñaron un sistema alternativo de sonido envolvente mejorando cualitativamente a todos los sistemas que había en esa época. El nombre elegido para dicha tecnología sería ‘Ambisonics’, que proviene del verbo latino ambire (ir alrededor, rodear) y el sustantivo sonus (sonido).

Una descripción menos técnica del sistema Ambisonic se trata de simular una esfera de 360 grados en donde el sonido proviene de cualquier dirección teniendo como referencia un punto central, el cual en este caso sería el micrófono de grabación Ambisonic para su posterior reproducción en un sistema estéreo.

Para poder registrar una escena sonora tridimensional al completo se emplean un total de cuatro micrófonos, el primero con formato omnidireccional y otros tres bidireccionales. El primero de estos mencionados registra la presión (W) y el resto registran el gradiente de la presión en cada uno de los ejes (X , Y , Z). Al almacenar cada uno de estos cuatro valores en una pista diferente, obtendremos el formato Ambisonics de primer orden, conocido como Ambisonics-B o B-format. Si comparamos con los formatos más comunes estéreo, la reproducción de estos cuatro canales no se limita a una simple asignación canal-altavoz, sino que el equipo reproductor ha de saber en qué posición se localizan cada uno de los altavoces para así decidir qué reproducir en cada uno de ellos, para que el oyente final que se encuentra en un punto determinado, los cuatro parámetros mencionados antes, presión y gradientes de velocidad, coincidan con los registrados. En la figura 1 se muestra cómo se realiza la decodificación de una señal sonora ambisonics.

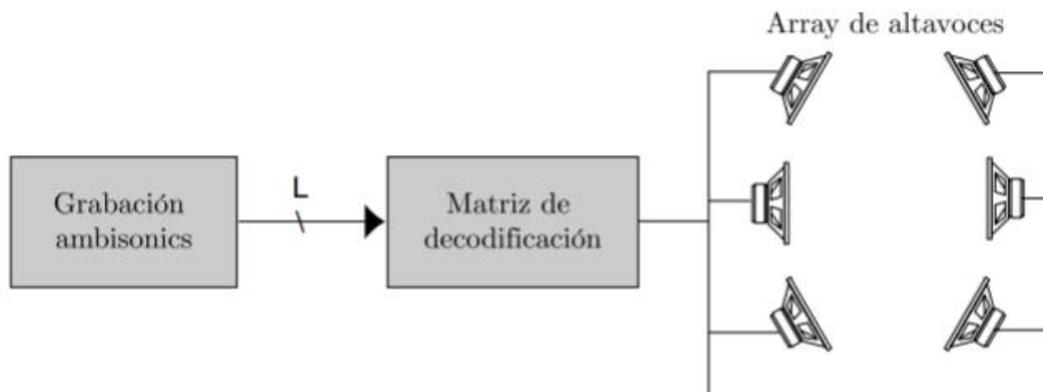


Figura 1. Decodificación de una escena sonora ambisonics.

Dejando de lado su sólida base técnica y sus innumerables ventajas, esta tecnología no había sido altamente aceptada comercialmente hasta hace poco, sobreviviendo en apenas unas pocas aplicaciones y los amantes de la grabación. No ha sido hasta la llegada de la tecnología tridimensional y la realidad virtual (tecnologías que requieren soluciones de audio 360°) cuando Ambisonics no ha sido aceptado y adoptado comercialmente.

La tecnología Ambisonics está formada por 4 fases [1]:

- Primera: Grabación con microfónica
- Segunda: Codificación de la señal en determinados formatos
- Tercera: Decodificación en canales específicos
- Cuarta: Reproducción de la señal a través de altavoces estéreos.

1.1 Métodos de grabación

1.1.1 Introducción a la metodología de grabación

[3]La representación Ambisónica se basa en lo esférico, siendo esta una descomposición armónica del campo sonoro, la cual proviene de escribir la ecuación de onda en el sistema esférico de coordenadas donde un punto \vec{r} , con un radio r , ángulo θ y una elevación δ como se muestra en la Figura 2.

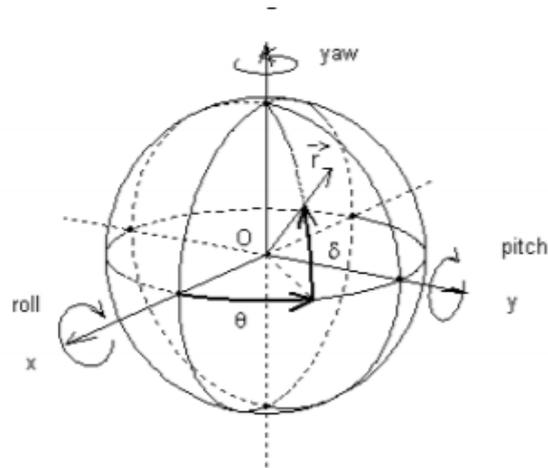


Figura 2. Sistema de coordenadas esféricas, con los tres grados de rotación elemental.

Matemáticamente hablando en cuanto a las señales usadas por Ambisonics, hablamos de la teoría de los armónicos esféricos. Esta teoría es una analogía de las series de Fourier, que son funciones armónicas capaces de reconstruir una señal gracias a la suma de sus componentes de manera individual.

En la teoría de Ambisonics el orden equivale al número de armónicos de la señal, siendo cero W la función omnidireccional.

- Para el primer orden, se incluye el orden cero junto con las señales de entrada X, Y, Z.
- El segundo; incluye los cuatro armónicos mencionados anteriormente junto con cinco armónicos más y así ininterrumpidamente para los siguientes órdenes.

El orden de trabajo para este proyecto será el primero (véase en la Figura 3).

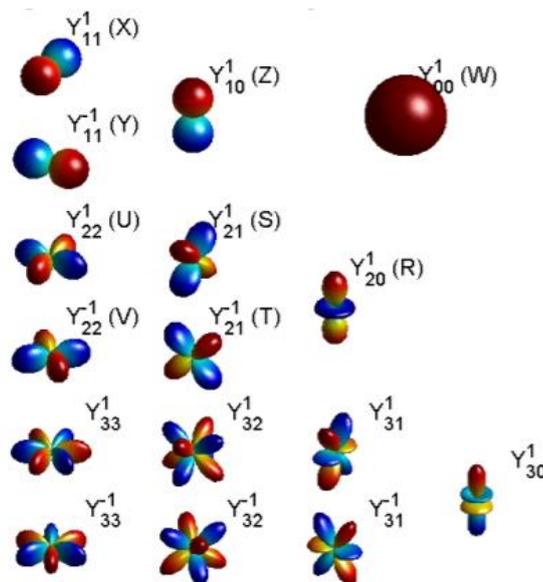


Figura 3. Vista 3D de funciones armónicas esféricas con habitual designación de componentes ambisónicos asociados.

1.1.2 Tipos de Microfonía y técnicas de grabación

En este apartado se van a reflejar las técnicas usadas para la grabación junto con los formatos de codificación que existen en Ambisonics.

[1] Una vez introducida la teoría en la que se fundamenta la tecnología Ambisonic, podemos decir que existen dos técnicas generales de grabación microfónica con las cuales obtener un registro adecuado para la codificación propia de Ambisonics.

En la actualidad, existen técnicas microfónicas para emular la manera con la que el oído humano recibe el sonido. Estas captan la señal mediante varios micrófonos, los cuales registran las cortas diferencias de tiempo, reflexiones y ecualización, de manera que se hace posible una futura reproducción en auriculares con un resultado estéreo similar con la escucha binaural que nos caracteriza.

La técnica microfónica regresa a los conceptos de origen marcados de Alan Blumlein, el cual solo añade dos micrófonos a la pareja de micros bidireccionales que forman el doble ocho de la técnica Blumlein:

- Uno bidireccional con orientación vertical
- Uno omnidireccional

Otra técnica diferente expone que colocando la pareja de micros que se sitúan en el plano paralelo a 90° y no a 45° se consigue obtener la variación de fase junto con la variación de intensidad de la señal recogida en cada uno de los micrófonos bidireccionales y un registro global con el micrófono omnidireccional. Utilizando esta técnica se recoge toda la información y variación del espacio tridimensional. Una vez codificada toda la información acústica, se permitirá realizar la decodificación de la manera que el usuario necesite.

Para solucionar la problemática por la cual era complicado hacer las cápsulas de los micrófonos coincidentes, se han creado micrófonos ambisonics. Actualmente existen dos micrófonos diseñados únicamente con este fin. Estos micros se denominan, TetraMic y soundfield.



Figura 4. Micrófono tipo TetraMic



Figura 5. Micrófono tipo Soundfield

Con la información descrita anteriormente, se concluye que existe dos técnicas mediante las cuales se puede obtener un registro adóneo para codificar en formato Ambisonic:

- Técnica Clásica.
- Técnica Ambisonic.

Técnica Clásica: La presente técnica, usa una composición de 4 micrófonos que coinciden entre sí. Como hemos mencionado anteriormente, tres de ellos bidireccionales, uno orientado a cada eje de coordenadas y el restante de formato omnidireccional utilizado para captar el ambiente. Para una efectividad máxima, se necesita que las cápsulas de los micrófonos sean 100% coincidentes, si no, afectaría al resultado negativamente igual que si se usan micrófonos independientes.

Técnica Ambisonics: Esta técnica avanzada, utiliza micrófonos orientados para solucionar la problemática existente con las grabaciones clásicas. Estos micrófonos contienen una matriz tetraédrica con 4 cápsulas normalmente cardioides. Las señales registradas de la cápsula dan 4 señales a la salida, en B Format.

Existen varios tipos como los ya mencionados Soundfield y TetraMic que fueron los primeros en salir al mercado. En la actualidad también existe más variedad como Sennheiser ambeo o Rode, el cual va a ser el utilizado para este proyecto.

Rode NT-SF1

Este dispositivo, es un micrófono ambisónico que captura un sonido envolvente de calidad de transmisión de 360° en A-Format en 4 pistas separadas que después serán convertidas en un nuevo set de 4 canales en B-format. Se trata de la primera colaboración entre la marca pionera mencionada antes SoundField (inventora del primer micrófono ambisónico comercial) y la marca de micrófonos Rode.



Figura 6. Micrófono RODE NT-SF1

Fabricado con precisión de latón, éste utiliza 4 cápsulas de micrófono cardioide de condensador dispuestas en una matriz tetraédrica para grabar todo el campo de sonido esférico. Cada sonido y cada dirección con una calidad de alto rango.

1.1.3 Configuración de Formatos. A y B

Actualmente y para las configuraciones a nivel microfónico Ambisonics, ofrece un resultado de cuatro señales las cuales cada una está relacionada entre sí. Cada formato obtenido diferente viene dado con el uso de las técnicas mencionadas anteriormente:

- Formato-A: Resultado de la configuración microfónica Ambisonics.
- Formato-B: Resultado de la configuración microfónica clásica. Identificando cada uno de los canales con las siglas W, X, Y, Z.

Formato-A

Con el uso de los micrófonos tetraédricos para realizar una grabación, se obtienen, como ya hemos mencionado antes, 4 señales en Formato-A. Dado que para trabajar en Ambisonics, se utiliza el Formato- B (las señales que se obtienen al utilizar una configuración microfónica clásica), será necesario realizar la conversión de Formato-A a Formato-B de las señales para poder trabajar adecuadamente.

Este tipo de micrófono utiliza un total de 4 cápsulas cardioides o subcardioides colocadas forma tetraédrica. Primero, se etiquetan las señales de las cuatro cápsulas, empleando el sistema de referencia cartesiano antropométrico, como se muestra a continuación

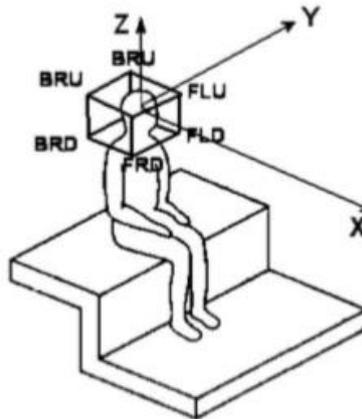


Figura 7. Posicionamiento de las cápsulas en cada eje

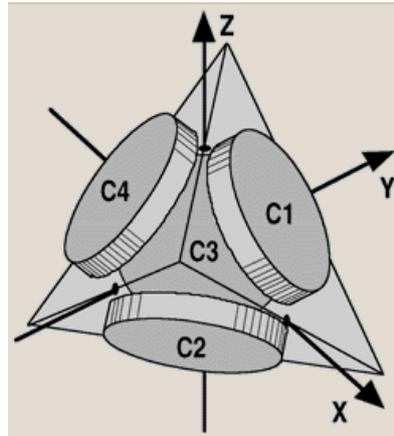


Figura 8. Posición de las cápsulas en X, Y y Z

Cada una de estas cápsulas se comparan con 4 de los 8 vértices de un cubo que rodean la forma de cabeza del oyente virtual, que imitan al sistema referente cartesiano. Estas caras se denominan de la siguiente manera:

Front-back
 Left-Right
 Up-Down

Cada uno de los vértices del cubo imaginario converge en una intersección que comparte tres lados de este cubo, y así, un micrófono con orientación hacia adelante, izquierda y arriba se etiqueta como FLU (Front-Left-Up) Si tomamos como referencia sólo el plano horizontal, se ordenan las señales de la siguiente manera:

FL: Front-Left
 FR: Front-Right
 BL: Back-Left
 BR: Back-Right

La transición A to B comienza con la grabación de los canales que se han recibido de la cápsula anteriormente mencionada sobre los canales W, X, Y y Z.

Las señales obtenidas se definen de manera matemática como las señales de 4 micros distintos que son coincidentes. La señal W corresponde a un micrófono omnidireccional, mientras que X, Y y Z son micrófonos de velocidad, mostrando un patrón de directividad que se asemeja a una figura de 8 o infinito.

Las operaciones matriciales realizan la conversión de Formato-A a Formato-B marcando los fundamentos de conversión. A continuación, se muestra la conversión matricial en la que se basa esa conversión:

$$\begin{aligned}W' &: FLU + FRD + BLD + BRU \\X' &: FLU + FRD - BLD - BRU \\Y' &: FLU - FRD + BLD - BRU \\Z' &: FLU - FRU - BLD + BRU\end{aligned}$$

Las cuatro señales obtenidas gracias a las operaciones mencionadas en el párrafo anterior no equivalen a las señales de salida para el estándar de Formato-B, ya que, requieren operaciones para ser filtradas. A continuación, se muestran dos filtros utilizados, aplicándose cada uno de ellos en función del canal de entrada.

Filtro Fw Filtro Fxyz

Filtro Fw

Cuando se aplica el Filtro Fw a W', tenemos como resultado W.

$$F_W = \frac{1 + \frac{j\omega r}{c} - \frac{1}{3} \left(\frac{\omega r}{c}\right)^2}{1 + \frac{1}{3} \left(\frac{j\omega r}{c}\right)}$$

Siendo:

r = distancia de cada cápsula al centro del tetraedro en metros
w = frecuencia angular en rad/s ($w = 2\pi f$)
c = velocidad del sonido en m/s (340 m/s)

Filtro Fxyz

Cuando aplicamos el Filtro Fxyz a cualquiera de los canales X', Y', Z' obtenemos como resultado X, Y, Z.

$$F_{XYZ} = \sqrt{6} \frac{1 + \frac{1}{3} \left(\frac{j\omega r}{c}\right) - \frac{1}{3} \left(\frac{\omega r}{c}\right)^2}{1 + \frac{1}{3} \left(\frac{j\omega r}{c}\right)}$$

Siendo:

r = distancia de cada cápsula al centro del tetraedro en metros
w = frecuencia angular en rad/s ($w = 2\pi f$)
c = velocidad del sonido en m/s (340 m/s)

Aunque afectan de manera leve a la fase, los filtros mostrados en las imágenes anteriores conservan la ecualización ganancial.

Formato-B

Las salidas, 4 señales correspondientes con el resultado de la grabación microfónica clásica explicada anteriormente. EL resultado es cada uno de los canales que obtenemos, que corresponde a la señal modelo en Ambisonics (X, Y, Z y W).

Podemos codificar la posición de un sonido mediante el estándar de cuatro canales del Formato B mencionado anteriormente dentro de un campo sonoro tridimensional:

$$\begin{aligned}X &= \cos A \cdot \cos B \text{ (frente-atrás)} \\Y &= \sin A \cdot \cos B \text{ (izquierda-derecha)} \\Z &= \sin B \text{ (arriba-abajo)} \\W &= 0.707 \text{ (señal de presión)}\end{aligned}$$

Siendo A el ángulo horizontal y B es el ángulo de la elevación. Limitando la localización de los sonidos en la esfera-unidad. Si se limita la posición de los sonidos a una esfera-unidad, garantizando que:

$$(x^2 + y^2 + z^2)$$

Siempre que sea ≤ 1 , de esta forma las ecuaciones se representan de la siguiente manera:

$$\begin{aligned}X &= x \\Y &= y \\Z &= z \\W &= 0.707\end{aligned}$$

Donde las primeras tres variables son las coordenadas fuente de sonido, W tiene un valor de 0.707 siendo distinto de 1 para que la distribución sea mucho más similar del nivel y los 4 canales. De la siguiente manera mostrada conseguimos una solución que mantiene el nivel. Haciendo que W varíe como se muestra en la siguiente solución, mantendrá el nivel total constante de manera adecuada:

$$W = 1 - 0,293 \cdot (x^2 + y^2 + z^2)$$

FuMa y **Ambix** son las conversiones ligadas al B-Format mencionado anteriormente. Estas conversiones de Ambisonics son muy parecidas, una de las diferencias se encuentra en la disposición secuencial de cada uno de los cuatro canales; uno se dispone en WYZX en vez de WXYZ.

1.2 F. Codificación

[2] Las señales estándar que vienen o se han generado en Formato-B tienen una distribución en la superficie de una esfera ‘unidad’ o dentro de ella. Los sonidos que no pertenecen a la esfera unidad no podrán decodificarse de manera válida y su situación cambiará, de manera que cuando el radio máximo sea = 1, su ubicación será la de la salida más cercana. Por esto, la distribución o coordenadas de la salida de sonido se ajusta a la siguiente ecuación:

$$(x^2 + y^2 + z^2) \leq 1$$

Donde ‘x’ hace referencia al recorrido en distancia del eje X, el parámetro ‘y’ corresponde al recorrido en distancia del eje Y y por último z corresponde al recorrido en distancia del eje Z. A continuación, se muestra la distribución de los diferentes ejes:

X: Frente-Atrás

Y: Izquierda-Derecha

Z: Arriba-Abajo

[2] Las coordenadas de una señal monofónica que se distribuye en la superficie del campo son las siguientes, siendo A en ángulo horizontal y B el ángulo vertical:

$$x = \cos A \cdot \cos B$$

$$y = \sin A \cdot \cos B$$

$$z = \sin B$$

Estas coordenadas son los multiplicadores que se utilizan para generar las señales en el B-Format. Estos multiplicadores se representan de la siguiente manera:

$$X = \text{input} \cdot \text{signal} \cdot x$$

$$Y = \text{input} \cdot \text{signal} \cdot y$$

$$Z = \text{input} \cdot \text{signal} \cdot z$$

$$W = \text{input} \cdot \text{signal} \cdot 0.707$$

Como hemos informado antes, el dato 0,707 se utiliza como multiplicador en la coordenada W lo que ayuda a obtener una distribución uniforme en la señal para cada uno de los cuatro canales.

1.2.1 Señales Ambisonics, Estéreo y Formato UHJ

Existe una incompatibilidad exhausta entre las señales en Formato-B y las señales de un sistema estéreo. Sin embargo, se puede combinar los parámetros X, W e Y que son necesarios para el

trabajo horizontal de una forma para conseguir una señal concordante con el estéreo y a su vez utilizando un proceso de decodificación idóneo podemos obtener la espacialización adecuada y así poder reproducirlo por este sistema.

[4] Este método conocido como UHJ, es utilizado para producir todo tipo de audiovisuales tales como emisiones, grabaciones, cintas, etc, compatibles con Ambisonics. Utilizando la codificación de en dos canales de X, Y y W como se muestra en la siguiente transformada logramos producir audiovisuales compatibles con ambisonics:

$$\begin{aligned}\text{Left} &= (0.0928 + 0.255j)X + (0.4699 - 0.171j)W + (0.3277)Y \\ \text{Right} &= (0.0928 - 0.255j)X + (0.4699 + 0.171j)W - (0.3277)Y\end{aligned}$$

La parte imaginaria que podemos ver en la anterior transformada nos muestra el desplazamiento de fase 90° a lo largo de la banda completa de espectro del audio respecto a la versión normal. Esto sucede cuando las señales pasan a través de su par de dispositivos de desplazamiento de fase 90° de banda ancha para codificar y decodificar.

1.3 F. Decodificación

En un principio, decodificar una señal una vez codificada en formato Ambisonic puede ser complicada en cuanto a la optimización de decodificadores para sistemas británicos, ya que utilizan técnicas psico acústicas, aunque no se utilicen mucho para áreas de gran dimensión debido a su poca productividad.

La decodificación Ambisonics se ajusta a dos fases:

- 1ª Fase: se reconstruye una aproximación de las señales registradas en Formato-B originales de la salida generada del formato UHJ. Cuantos más canales UHJ sean utilizados, más precisa será la aproximación de la decodificación al Formato-B.
- 2ª Fase: Según las direcciones de procedencia de las señales, se aplican diversas curvas de ecualización. Este filtro imita de manera muy similar a la manera en la que los humanos con su sistema auditivo localizan los sonidos para frecuencias en diferentes rangos.

Las salidas del formato Ambisonics añaden la combinación de las señales resultantes de utilizar el filtro para obtener las señales en Formato- B que vienen dadas por su localización respecto al centro de la matriz

[2] A continuación se muestra cómo serían las señales de 4 esquinas planas en matriz cúbica. Planos U y D:

$$\begin{aligned}
 \text{LFU: } & W + 0.707(x + Y + Z) \\
 \text{RFU: } & W + 0.707(X - Y + Z) \\
 \text{LBU: } & W + 0.707(- X + Y + Z) \\
 \text{RBU: } & W + 0.707(- X - Y + Z) \\
 \text{LFD: } & W + 0.707(X + Y - Z) \\
 \text{RFD: } & W + 0.707(X - Y - Z) \\
 \text{LBD: } & W + 0.707(- X + Y - Z) \\
 \text{RBD: } & W + 0.707(- X - Y - Z)
 \end{aligned}$$

Para cada altavoz, el multiplicador 0.707 ofrece una respuesta direccional (cardioide). Esta configuración es la óptima para un espectador cercano a los altavoces o fuera de la matriz de altavoces. En el caso que el receptor se encuentre dentro de la matriz, se recomienda utilizar una manera distinta la forma hipercardioide que nos llevará a un aumento del multiplicador mejorando la imagen de los espectadores más céntricos.

1.4 Física de la audición binaural.

La raza humana escucha los sonidos con el oído, que consta de tres partes, empezando por el oído externo donde nos encontramos el pabellón auditivo, después el oído medio donde se encuentran partes como los huesecillos y por último el oído interno donde posteriormente transformar la señal e interpretarla en el cerebro. El ser humano es capaz de recibir señales con frecuencias entre los 20 Hz y 20kHz.

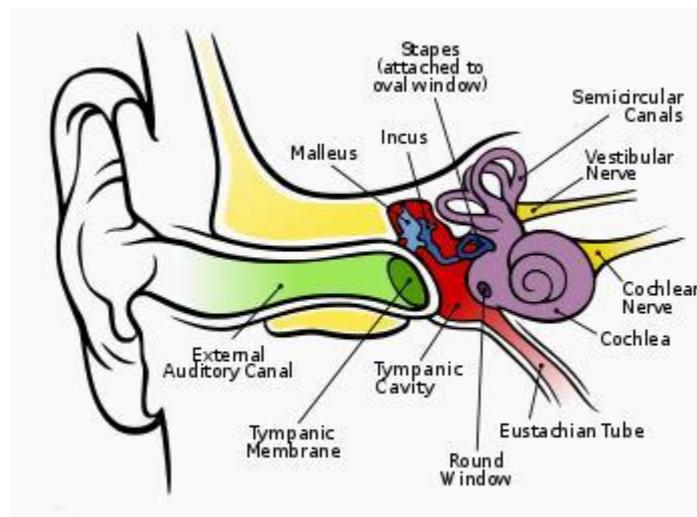


Figura 9. Órgano Auditivo Humano

[5] El tímpano que pertenece al oído externo es el encargado de diferencias los cambios de presión. Después modifica esta señal de acústica a mecánica para ser transferida al oído medio. El estribo genera unas ondas que provocan el movimiento en las células ciliadas que estimulan los nervios conectados al cerebro.

La distancia, azimut y elevación son los parámetros que nos llevan al ser humano a localizar la señal recibida de forma que tomando las coordenadas esféricas situamos la procedencia de este.

1.4.1 Psicoacústica de la localización espacial.

1.4.1.1 Indicadores en acimut.

Lord Rayleigh (físico y catedrático británico), tras diversos estudios, hizo un desarrollo conocido como la Teoría Dúplex en 1907. Estos estudios se relacionaron con experimentos con tonos puros [7], en los cuales se especificó de forma explícita la forma de localizar los sonidos en el plano conocido como horizontal o azimut por medio de dos parámetros conocidos como interaurales:

- diferencia de tiempos interaural ITD (*interaural time difference*).
- diferencia de nivel interaural IID (*interaural level diferente*).

Una ITD se define coloquialmente como la diferencia de tiempo que tarda la onda en llegar a un oído u a otro. Para el cálculo de este parámetro Rayleigh definió una sencilla fórmula para definir la distancia que debe viajar la onda hasta el oído más lejano: este recorrido se planteó de con la siguiente fórmula: **$a(\text{ángulo azimut } (\theta) + \text{sen}\theta)$**

Como representación de lo explicado, se muestra la Figura 10 que hace referencia a cómo es el cálculo de la diferencia temporal entre un oído u otro siendo una onda plana y una cabeza esférica.

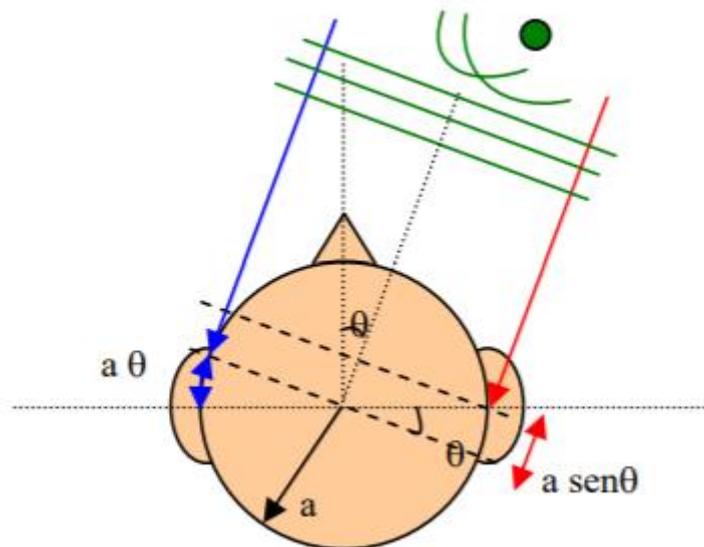


Figura 10. Representación del cálculo esquemática de ITD

Tras los estudios en ondas incidentes llevados a cabo por Rayleingh para desarrollar una solución de la ecuación de onda en donde demostró la manera en la que una onda plana puede llegar a difractarse por una esfera de similitud a la cabeza humana o una esfera rígida, definió en paralelo a sus estudios otro parámetro mediante el cual existía una diferencia de amplitudes para cada uno de los sistemas auditivos. Este parámetro es conocido como ILD que se muestra en la Figura 13.

En este caso, el parámetro ITD no es perfecto y tiene limitaciones. Por una parte, hay varios puntos que según donde esté colocada la fuente de sonido se provoca un valor exacto de ITD. Como se muestra en el ejemplo a continuación, una fuente que genera una emisión de señal a 45° provoca el mismo valor de ITD que otra fuente con una localización a 135° del objetivo. Cuando se da esta coincidencia en la emisión de señal o sonido, se le conoce como confusión delante-atrás.

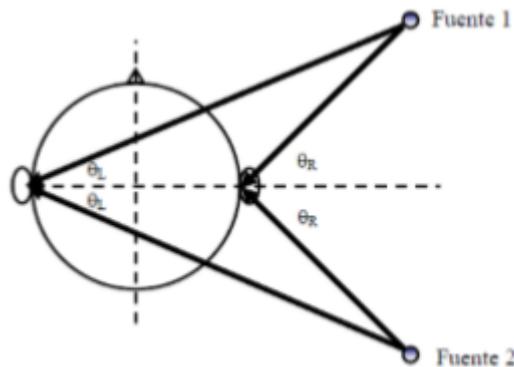


Figura 11. Limitación ITD confusión delante-atrás.

La diferencia de fases marca cómo se percibe la diferencia de tiempos interaurales. En este caso, por debajo de los de los 1.6 kHz, tenemos escasa información proporcionada por la ITD. Sin embargo, a frecuencias más altas de 1.6 kHz se produce confusión dado que provoca varios períodos de desplazamiento. Desde de la mencionada frecuencia en adelante, la señal sinusoidal generada puede girar ≥ 1 long. onda, dando como resultado que el oído humano capte ITDs de menor valor.

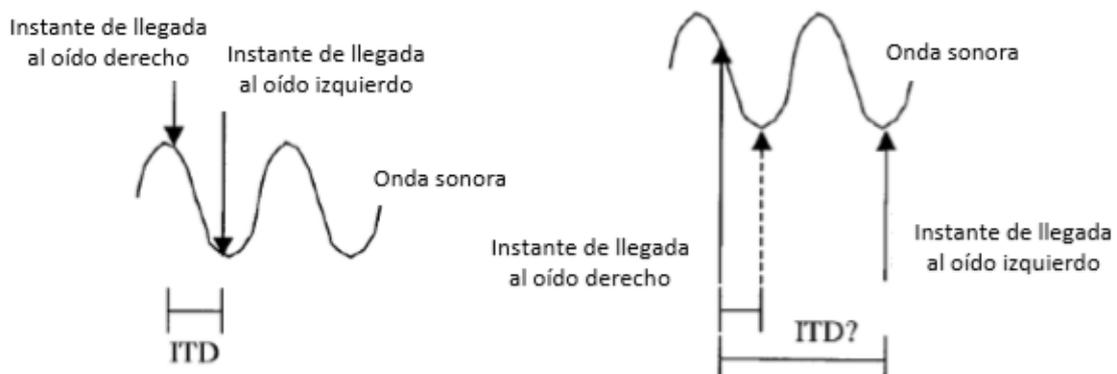


Figura 12. Izq: ITD más pequeño a 1 longitud de onda \rightarrow un valor de azimut.
 Dr: ITD más grande que 1 longitud de onda $\rightarrow \geq 1$ valores de azimut.

[6] En contra, teniendo valores más altos que 1.6 kHz, es el IID el que domina el sistema auditivo debido a que las características que proporcionan tanto la cabeza y la longitud de onda produce efecto por el cual el oído más lejano recibe menos nivel de señal. En la Figura 13 se muestra esta teoría de forma visual:

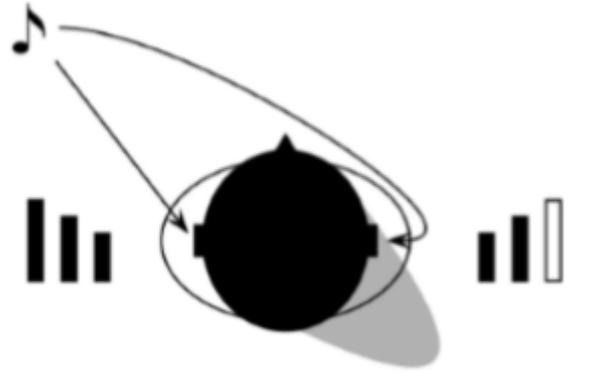


Figura 13. Efecto de difracción de la cabeza y la longitud de onda sobre IID.

En resumen, podemos afirmar que los parámetros explicados ITD e IID son complementarios en cuanto a la determinación del ángulo de incidencia y dependen de la frecuencia en su totalidad y que la ITD por encima de 1.6 khz ofrece un desfase de las señales envolventes y por debajo de 1.6 khz un desfase de las señales portadoras.

1.4.2 Indicadores de elevación y HRTF

[7] Siendo diferente al mecanismo descrito en el apartado anterior sobre cómo se sitúan las señales el plano azimut, en el plano medio la raza humana utiliza diferentes sistemas para localizar los ángulos de elevación. Existe una teoría común que es la más aceptada en la sociedad hace referencia a las diferencias de timbre. La manera en la que se logra esto y, existiendo diferentes teorías al respecto, la más aceptada dentro del campo de la acústica se refiere a las diferencias de timbre. Los cambios de timbre se producen cuando varía el ángulo de la señal incidente.

A su vez se puede afirmar que la percepción de la localización del sonido en el plano medio depende de cada ser humano, siendo cabeza, hombros y oídos factores que dependen de la función de transferencia de cada persona y para cada sentido direccional de procedencia del sonido. Esta función, la cual es distinta para cada ser humano, con parámetros distintos se denomina HRTF (Head-Related Transfer Functions).

Para el desarrollo del presente proyecto se utilizarán valores medios estimados, sacados del cálculo de los parámetros HRTF de un individuo. Los plugins utilizados ya tienen intrínsecos estos valores para poder trabajar con ellos.

La manera de calcular las HRTF se puede realizar de distintas formas, equipamiento distinto, análisis distinto, estudios o laboratorios distintos etc. Para ello existe una función general con diferentes parámetros donde el ángulo de elevación es θ , el ángulo de azimut es φ , la distancia radial es r y la frecuencia f . En la figura 14 se muestra la fórmula.

$$HRTF(r, \theta, \varphi, f) = \frac{P(r, \theta, \varphi, f)}{P_s(f)}$$

Figura 14. Función HRTF

1.4.3 Indicadores de distancia y vector distancia.

Para finalizar con esta pequeña introducción a los principios binaurales vamos a definir el vector distancia en cuanto a distancias cercanas que vienen siendo entre 3 y 15 metros y distancias lejanas a partir de 15 metros.

El nivel de presión sonora de la señal acústica en distancias cercanas determina el vector distancia. En espacios abiertos, la teoría nos dice que la señal disminuye 6dB siempre que se duplique la distancia de la fuente de sonido. Para las distancias cercanas ente 3 y 15 metros el sistema auditivo del ser humano es bastante eficiente a la hora de saber localizar en distancia la procedencia de la señal.

Sin embargo, para las distancias mayores a 15 metros el aire juega un papel muy importante ya que modifica el espectro y el nivel de intensidad. Las frecuencias altas son atenuadas por el efecto del aire y teniendo en cuenta lo dicho, el ser humano tiene mayores problemas a la hora de situar la fuente de sonido a distancias más lejanas.

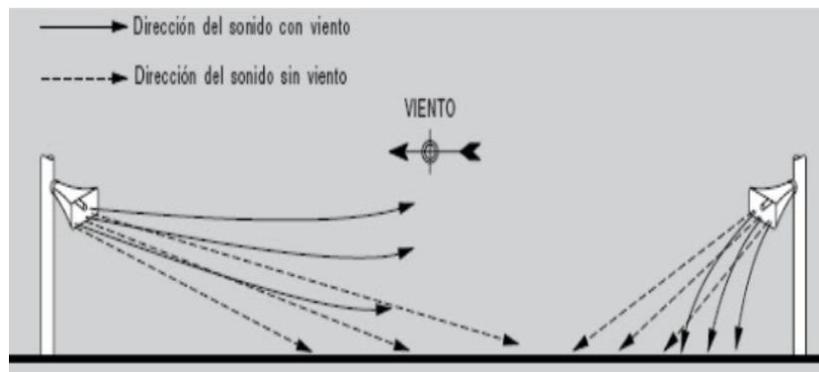


Figura 15. Efecto del viento sobre el sonido.

1.5 Respuestas binaurales

En la actualidad hay dos métodos de respuestas binaurales: la primera la respuesta al impulso relacionadas con la cabeza (*head-related impulse response, HRIR*) y las respuestas al impulso binaurales de sala (*binaural room impulse response, BRIR*). Ambas respuestas se graban por pares (cada uno para cada conducto auditivo), una pareja de cada respuesta determina la sensación sonora de un punto en el espacio.

[10] Conocemos las respuestas HRIR como el impulso binaural resultante de los cambios que se producen en el sonido por medio del cuerpo humano antes de llegar al órgano reproductor. Para realizar la grabación de las HRIR son grabadas en una cámara específica (anecoica) en la cual se emplea una cabeza especial conocida como ‘cabeza binaural’, también vale una persona.

[11] Conocemos la respuesta BRIR como el impulso binaural de una habitación o lugar específico. Para obtener estas respuestas se utiliza un maniquí binaural o una persona en una localización con un nivel de reverberación adecuado.

1.6 Aplicaciones

La medición binaural, se utiliza para sintetizar los sonidos binaurales. Midiendo los dos tipos de respuesta mencionados antes, HRIR y BRIR de un humano o de un maniquí, somos capaces de obtener dichos parámetros al impulso que, realizando una convolución con cualquier señal o sonido, reproduce el mismo sonido intentando imitar una situación real como si esa misma persona o maniquí que nos ha ofrecido los parámetros, el sonido binauralizado. Realizando una medición muy minuciosa de los parámetros cada 5 grados, podría situarse el sonido en cualquier dirección o distancia, logrando desplazar el sonido en plano horizontal, plano medio y también en distancia.

1.7 Conversión a binaural

Una vez introducidos los principios físicos binaurales, vamos a explicar la conversión de las señales ambisónicas a señales binaurales.

A la hora de reconstruir una señal binaural nos encontramos con diferentes problemas. Uno de ellos es reproducir este sonido desde un punto de vista diferente. Siendo muy compleja la escena acústica que se quiere idear, el receptor va a captar de ella sólo lo que percibe en sus oídos.

La tecnología Ambisonics es utilizada en la industria virtual, ofreciendo experiencias en tres dimensiones para videojuegos y realidad virtual. Los auriculares en la mayoría de los usuarios son el medio final para escuchar los sonidos y para poder reproducir la experiencia final en estos auriculares, la industria utiliza la tecnología de audio binaural.

El sonido espacial en los auriculares es posible gracias a las tecnologías de audio binaural. En esencia, un procesador binaural recibe una entrada de audio y una dirección en la que posicionarlo. el procesador agrega señales auditivas a la señal, de modo que cuando se reproduce en los auriculares se experimenta en la posición virtual establecida.

[12] Decodificar los canales de las señales ambisonics teniendo en cuenta las salidas de audio es la manera en la cual se realiza el proceso para este tipo de señales. Tras esta decodificación de los canales los outputs son configurados hacia un procesador binaural que se limita a ubicar las señales como si un humano lo estuviera retransmitiendo, de esta manera logramos una experiencia casi real y utilizando unos auriculares logramos un sonido esférico inmersivo.

Capítulo 2. Hardware

Dado que el principal objetivo de este trabajo es grabar la respuesta ambisónica de conjuntos musicales acústicos, para ello se requieren diferentes elementos *hardware* que serán utilizados para llegar a tal fin.

2.1 Dispositivos

Infraestructura Hardware

- Ordenador portátil Asus
- Auriculares Shinnister
- Trípode para micrófono
- Micrófono Rode NT-SF1
- Interfaz de audio Roland Octa-capture UA 1010

2.1.1 *Micrófono Rode NT-SF1*

El micrófono que se va a utilizar en el desarrollo del presente proyecto se trata de un micrófono ambisónico que captura un sonido envolvente de calidad de transmisión de 360° en Formato-A en 4 pistas separadas. Se trata de la primera colaboración entre la marca pionera de ambisonic, SoundField, y la marca de micrófonos RODE (Fig 16).

Está fabricado en latón macizo y utiliza 4 cápsulas cardioides de condensador dispuestas en una matriz tetraédrica para grabar todo el campo de sonido esférico.

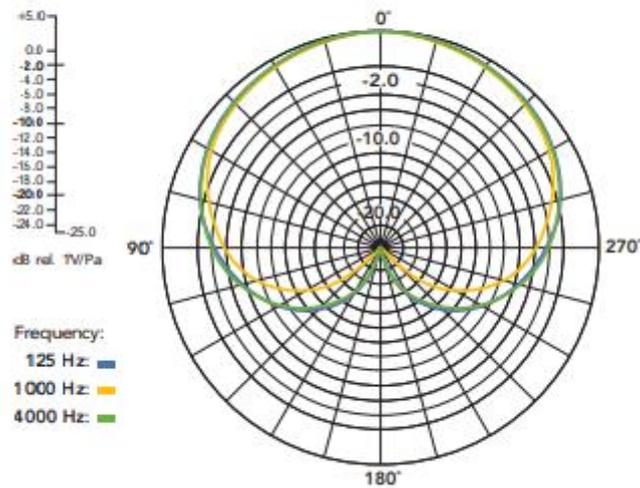
Está diseñado y desarrollado para crear audio envolvente para videojuegos, transmisiones deportivas, películas o contenido de calidad virtual (VR), este dispositivo es perfecto para toda grabación de sonido envolvente, y debido a la fácil conversión a Formato-B es capaz de ser manipulado en cualquier dirección en post-producción.



Figura 16. Micrófono Rode NT-SF1

A continuación, se muestra el patrón polar del micrófono de Rode. Como podemos observar, es un patrón cardioide. Su nombre viene por la forma de corazón del diagrama, que se traduce en que radian o captan la señal hacia o desde la parte frontal, teniendo un mínimo de sensibilidad en su parte posterior donde se produce una atenuación gradual.

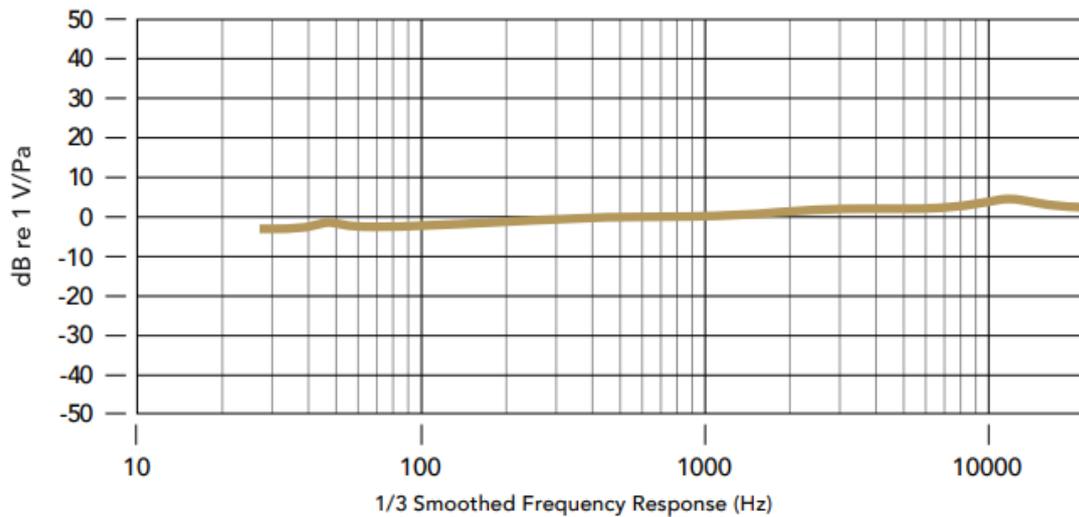
POLAR PATTERN



En cuanto a la respuesta en frecuencia que se muestra en la siguiente imagen, observamos la planitud de la señal en la mayoría del rango de frecuencias que el humano puede captar que va desde los 20 Hz hasta los 20000 Hz)

Esta es una de las principales razones por las cuales hemos elegido este modelo de micrófono ambisonics empleado en la medida de grupos acústicos.

FREQUENCY RESPONSE



2.1.2 Tarjeta de Audio Roland Octa-Capture. Modelo UA 1010

Para realizar la transmisión de las cuatro señales de entrada de cada uno de los micros que lleva nuestro dispositivo de grabación al activo portátil y desde ahí poder enviar las salidas a través de auriculares L&R, se necesita utilizar una interfaz de audio. Para ello, el departamento de audio me ha facilitado una interzar Roland Octa-Capture versión UA 1010, que se muestra en la figura 17.

Este dispositivo está diseñado para producciones de audio de alto nivel, combinando componentes de gran calidad, ingeniería con la garantía de la marca Roland junto con la tecnología propia de previos y streaming.



Figura 17. Interfaz de audio Rolando Octa-Capture UA 1010

Esta interfaz cuenta con 10 entradas y 10 salidas, ocho VS PREAMPS y una extraordinaria calidad de sonido de 24-bit/192Khz. El Octa-Capture define un nuevo estándar en interfaces de audio USB 2.0.

Capítulo 3. Software empleado en las medidas

Para el desarrollo del presente proyecto, grabación y procesado de las señales, se han empleado diversos programas detallados a continuación.

3.1 Aplicaciones

Infraestructura Software

- Reaper Digital Audio Workstation
- The Ambisonics Toolkit (ATK) plugins
- RØDE PLUGIN B to Ambix

3.1.1 *Software de procesado. Reaper Digital Audio.*

Reaper es el software mediante el cual vamos a trabajar para el desarrollo del presente proyecto. principal en el que se va a trabajar para el desarrollo del presente proyecto. Esta aplicación se utiliza para la producción de señales digitales en ordenadores, ofrece herramientas para la grabación, edición y procesado de señales. Es compatible con muchos hardwars distintos, formatos digitales y es capaz de ser programada y modificada al completo.

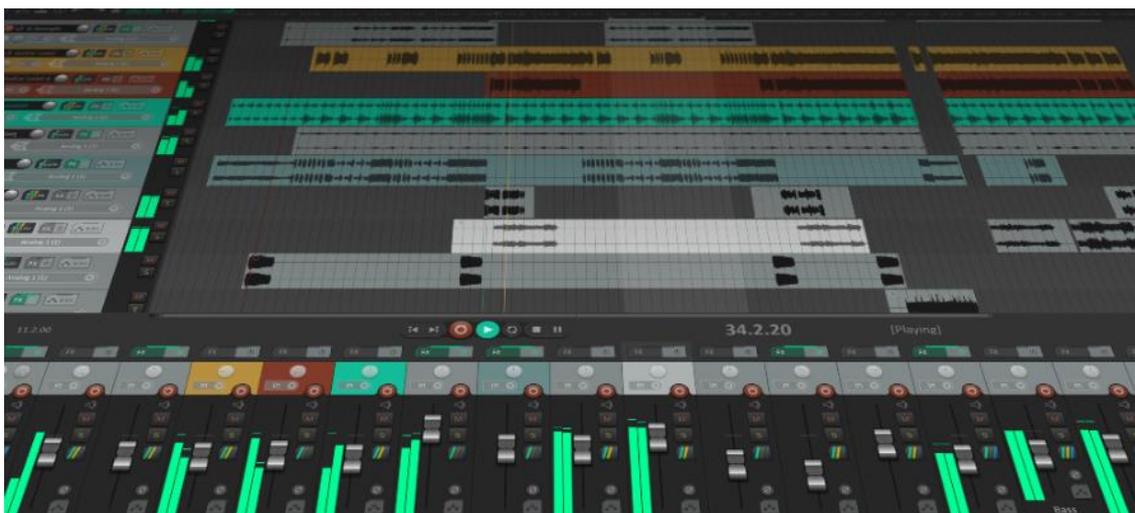


Figura 18. Reaper Digital

3.1.2 *Plugins ATK*

ATK engloba un conjunto de herramientas tanto clásicas como más novedosas para que el productor trabaje con el sonido envolvente ambisonics. Entre sus principales funcionales incluye codificaciones, transformadores, decodificadores y muchas de las herramientas que son necesarias para la realización del presente proyecto, en concreto codificadores y decodificadores.

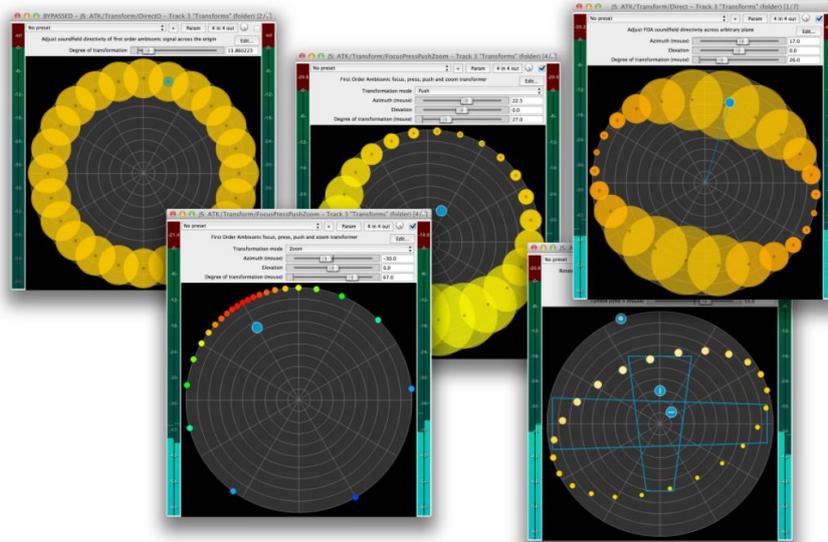


Figura 19. Plugins ATK

Existen en la actualidad diferentes formatos para el desarrollo de plugins de audio. A continuación, se explican los más habituales del sector.

- VST. Estándar diseñado y desarrollado por Steinberg para poder realizar la conexión de plugins de audio al software que actúa como host. Este formato es el más utilizado actualmente para el desarrollo de los plugins, siendo este el primer formato en ser gratuito.
- AU. Más conocido como Audio Unit, siendo este el estándar utilizado por la compañía Apple desarrollado para la arquitectura de plugins. Un formato equivalente al mencionado anteriormente siendo su estructura muy similar y muchas veces es común la compilación en ambos formatos. Estos plugins cuentan con su propia versión VST y AU.
- VST3. Presenta una actualización con nuevas características y mejoras funcionales con respecto al formato VST. Las principales mejoras son las siguientes: entradas y salidas dinámicas, adaptando el plugin a la pista o bus donde trabaje. Consigue una mayor activación y desactivación de buses en VTSs que cuenta con múltiples salidas; organización lógica de los parámetros en estructura de árbol; interfaz re-escalable; múltiples entradas y salidas MIDI; etc.

3.1.3 RØDE PLUGIN B to Ambix

[13] El complemento desarrollado por Rode, ofrece un nuevo enfoque para el procesamiento ambisónico, junto con su interfaz intuitiva. Ofrece una representación ambisónica precisa Su funcionamiento evita las matrices y los filtros de corrección y utiliza procesamiento de dominio de frecuencia de última generación que ofrece una precisión espacial incomparable en todas las frecuencias.



Figura 20. Plugin Rode para cambio a Formato B

Tras capturar el sonido en Formato A con una precisión y calidad muy alta, el complemento de Rode ofrece la espacialización y reproducción de sonido preciso.

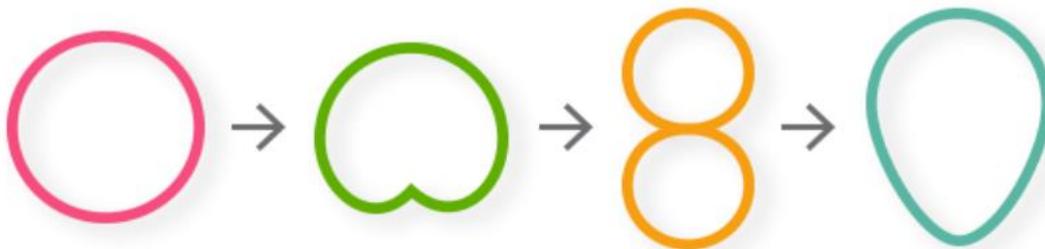


Figura 21. Cambio de patrones polares

Este plugin tiene la capacidad de cambiar el patrón polar de los micrófonos virtuales en producción. Permite al usuario crear todo, desde patrones omnidireccionales cardioides hasta creación de patrones tipo escopeta, utilizando desarrollos recientes en tecnología de formación de haces, que ofrece una nueva dimensión en la configuración y el control del sonido.

El complemento desarrollado por Rode permite al usuario configurar rápidamente mezclas para todos los formatos de sonido envolvente comunes, desde 5.1 hasta 7.1.4. Los formatos personalizados y las matrices de altavoces se pueden atender rápidamente, lo que brinda una flexibilidad y control incomparables.

Gracias a la investigación y el trabajo, se ha logrado una visualización única del nivel de sonido proporcionando información inmediata sobre los niveles y el paisaje sonoro, mientras que el modo 'avanzado' abre todos los parámetros y controles para lograr lo último en modelado de sonido.

Capítulo 4. Metodología y Pruebas realizadas.

4.1 Pruebas realizadas en laboratorio

Para las pruebas y medidas realizadas en el laboratorio antes de realizar las grabaciones a conjuntos acústicos musicales, se ha seguido el siguiente diagrama de bloques.

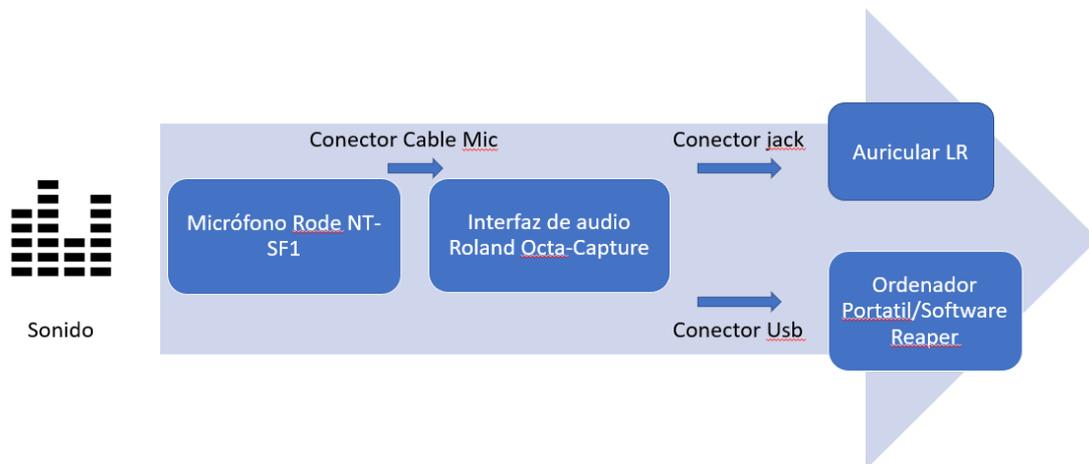


Figura 22. Diagrama de bloques

4.1.1 Procedimiento Experimental en laboratorio

En primer lugar, ha sido necesario aprender a manejar con soltura el software REAPER. Este software viene con una amplia variedad de personalización y efectos de producción de uso cotidiano. Este programa ofrece herramientas como ReaEQ, ReGate, ReaDelay etc. Su gran variedad de plugins intrínsecos van desde efectos estándar hasta aplicaciones específicas para dispositivos MIDI y audio. Reaper es compatible con todas las versiones VST lo que hace que se pueda utilizar con casi todos los complementos comerciables disponibles. También es utilizado para la edición de video.

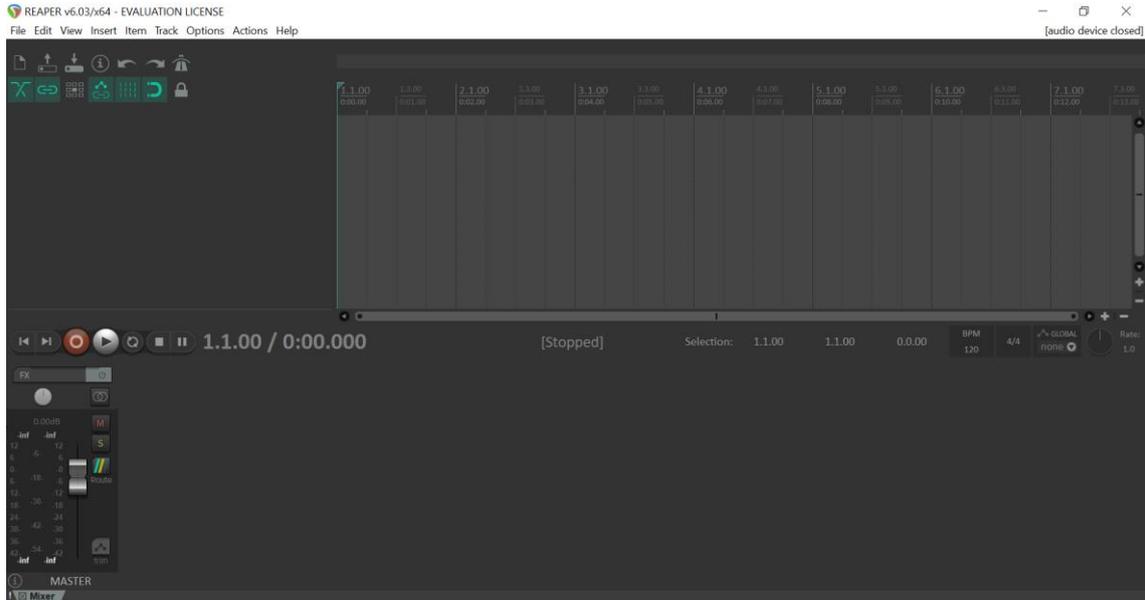


Figura 23. Pantalla inicio Reaper.

Para las primeras pruebas necesarias para conocer el software y el material necesario para la grabación y producción, la universidad me ha facilitado el laboratorio del instituto ITACA. En el laboratorio hemos comprobado el funcionamiento del software, y para ello se hicieron grabaciones de conversaciones cotidianas entre varias personas y pruebas de audio con una flauta dulce alrededor del micrófono. Estas grabaciones de prueba como a conjuntos acústicos deben estar grabadas en formato estéreo con entrada de 4 canales, siendo cada canal, cada uno de los micrófonos cardioides del micrófono Rode utilizado para el desarrollo del presente proyecto. A continuación, se muestra cómo queda una de las primeras grabaciones realizadas en el laboratorio de la flauta dulce alrededor del micrófono.

Para esta primera prueba, debemos de general una pista para la grabación, configurarla con un total de 4 canales de entrada que hacen referencia a cada uno de los micrófonos cardioides.

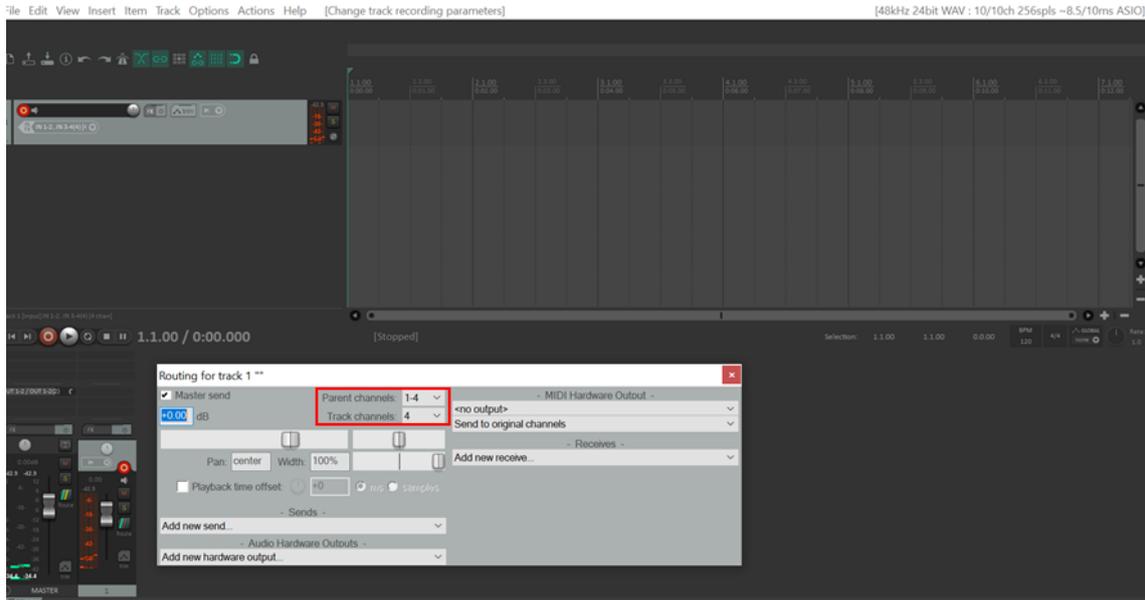


Figura 24. Configuración de 4 track channels desde Routing.

Tras las grabaciones de las primeras pruebas de laboratorio se obtiene una pista de cuatro canales con la señal recogida por cada uno de los micrófonos como se muestra en la figura 25.

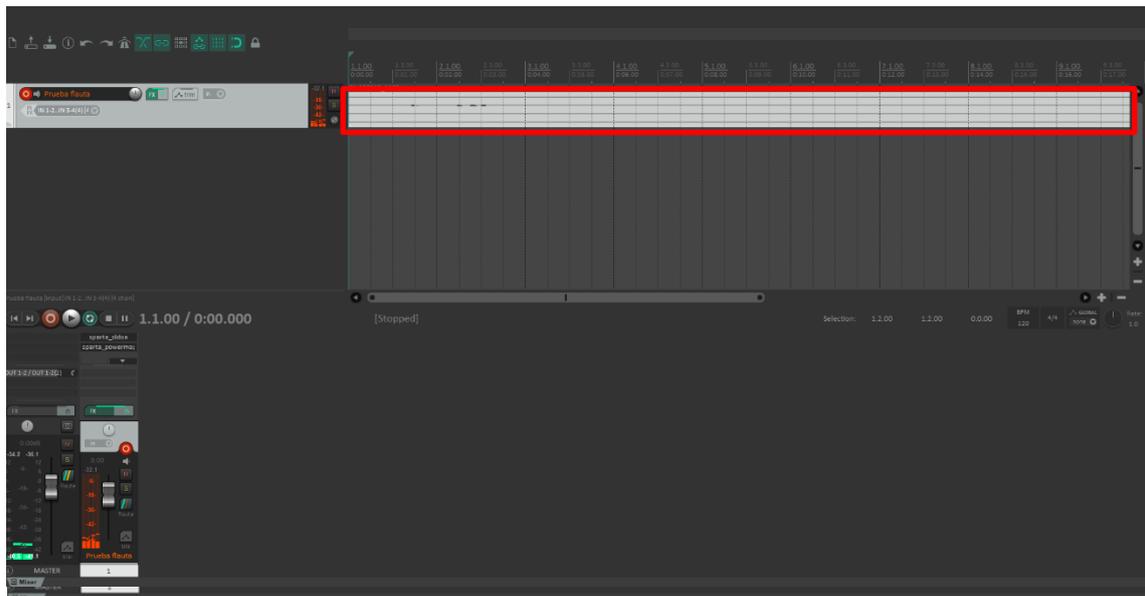


Figura 25. Primera prueba con flauta dulce.

Tras ir comprobando que las grabaciones tenían buena calidad e íbamos por el buen camino, pasamos al siguiente nivel y fuimos experimentando con los filtros en las grabaciones de prueba, tanto de la flauta como de las conversaciones entre nosotros. En primer lugar, tenemos que aplicar el filtro que la marca Rode ha desarrollado para la conversión entre formatos. En este caso se quiere obtener la señal de 4 canales en formato B. Para ello se aplica el filtro molstrado en la figura 26 que pasa la señal de formato A a formato B (Ambix).

Para abrir el menú y seleccionar el filtro que necesitamos se vamos a la opción Fx que nos ofrece el software y después buscamos el filtro que vamos a utilizar para el cambio de formato.

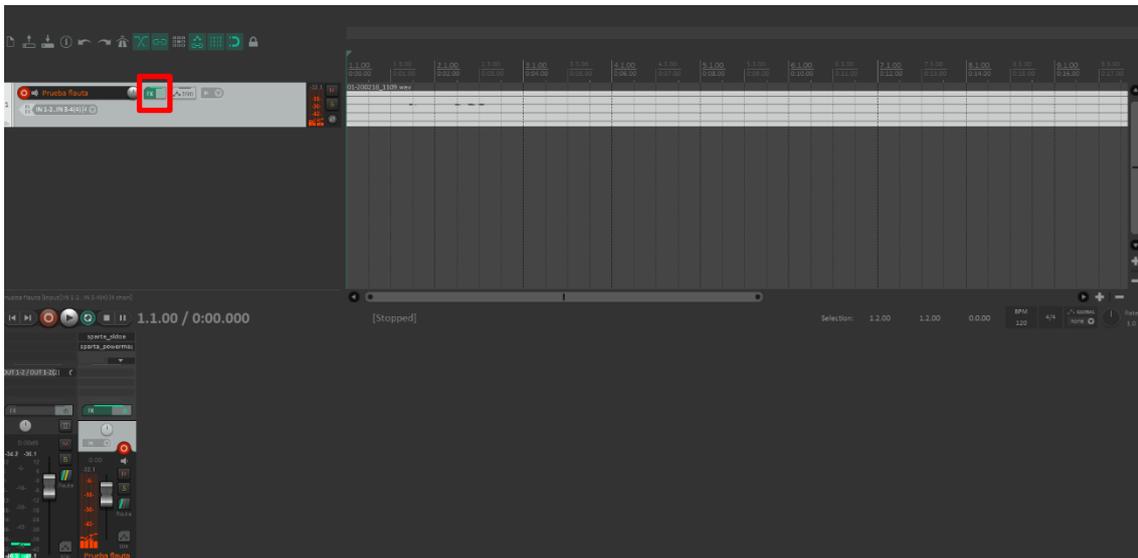


Figura 26. Opción de inserción de filtro.

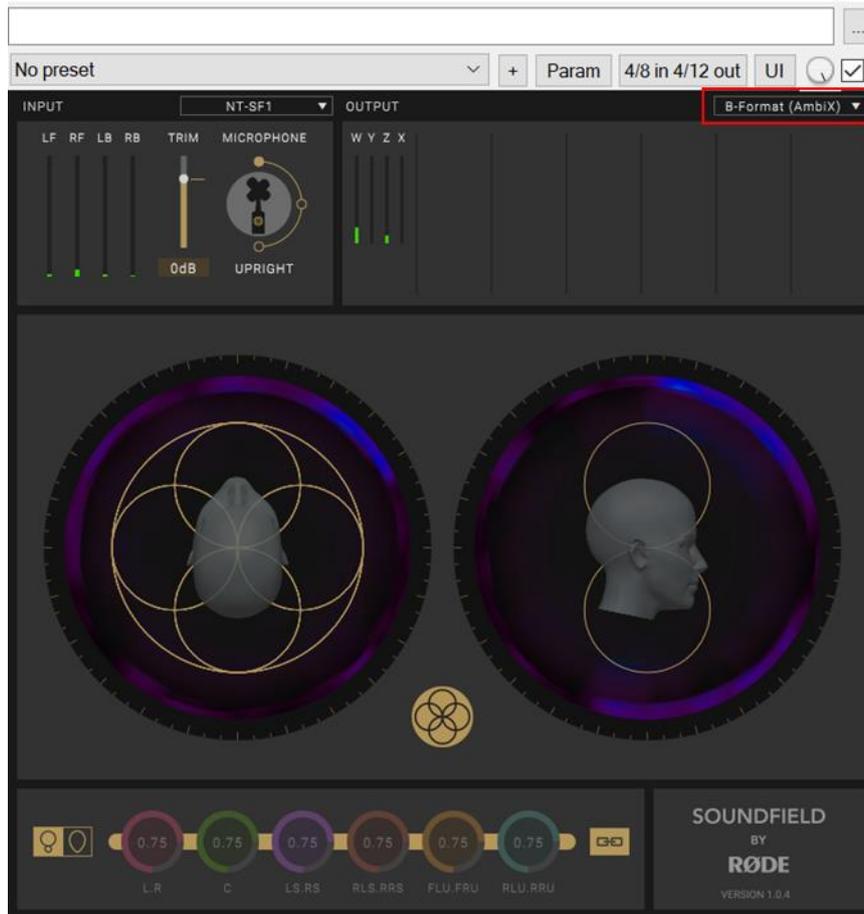


Figura 27. Filtro Rode. Formato A to Formato B.

Como observamos en los gráficos del filtro, tenemos 4 entradas: LF, RF, LB y RB que serán grabadas en una nueva pista en formato W, Y, Z y X.

Para poder guardar la salida del filtro con las nuevas condiciones en Formato B de cada uno de los canales grabados, hace falta generar una nueva pista que graba la salida del filtro. Para poder realizar una configuración óptima para grabar en esta nueva pista la salida del filtro hace falta realizar un enrutamiento desde la configuración Route, que se puede hacer de dos maneras distintas.

- La primera puede ser enviando desde la ruta de la primera grabación la salida de esta hacia la entrada de la nueva pista generada como se muestra en la figura 28 (Add new send).
- Por el contrario, desde la ruta de la pista nueva creada, grabar la salida de la primera grabación como se muestra en la figura 29 (Add new receive).

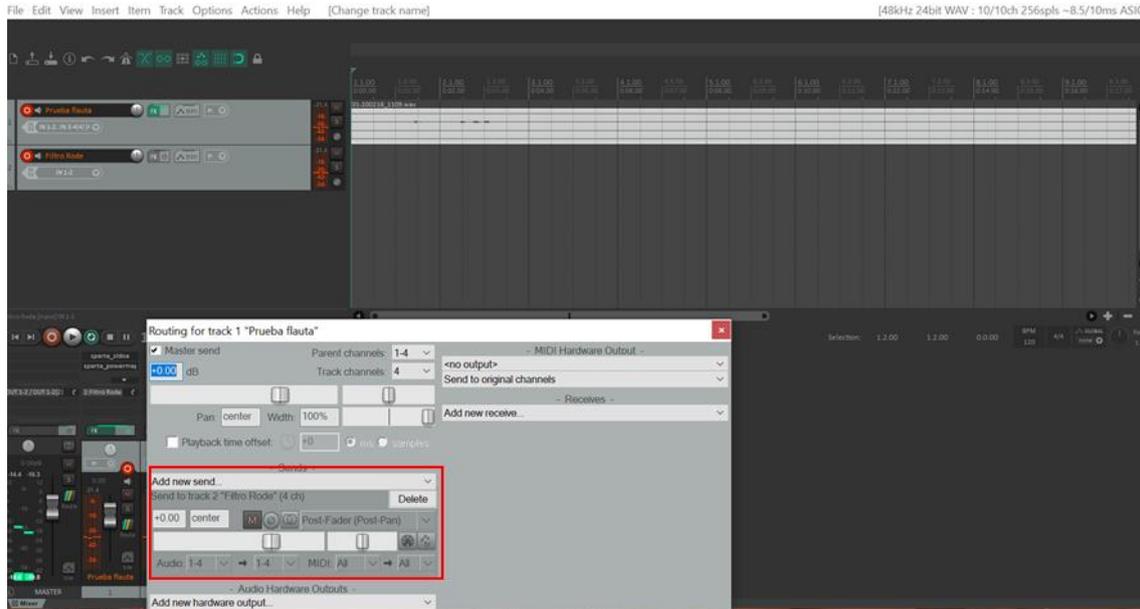


Figura 28. Configuración ruta desde primera pista.

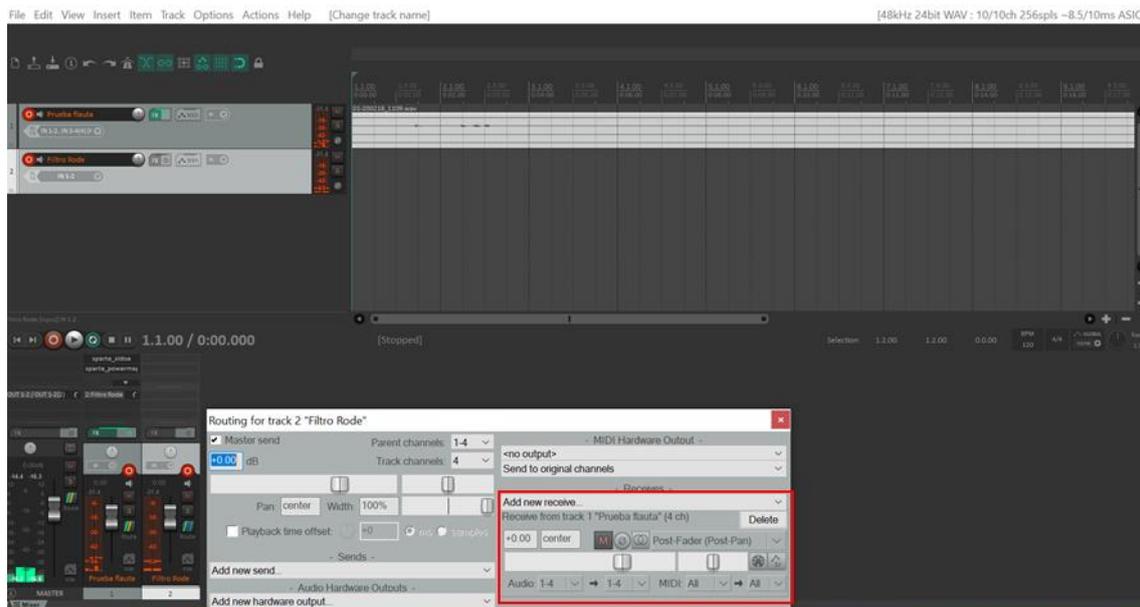


Figura 29. Configuración ruta desde nueva pista.

Tras la creación de la pista receptora de la señal filtrada por el filtro diseñado por Rode para el cambio de formato de A to B necesario para posteriormente aplicar el filtro adecuado para su transformación a formato binaural, se utiliza el método de grabación tradicional.

Como observamos en la figura 30 ahora tenemos una nueva pista sobre la que trabajar ya en formato B. La configuración del input en este caso debe ser nulo ya que la grabación viene sobre la salida de la primera pista tras pasar por el filtro anteriormente mencionado.

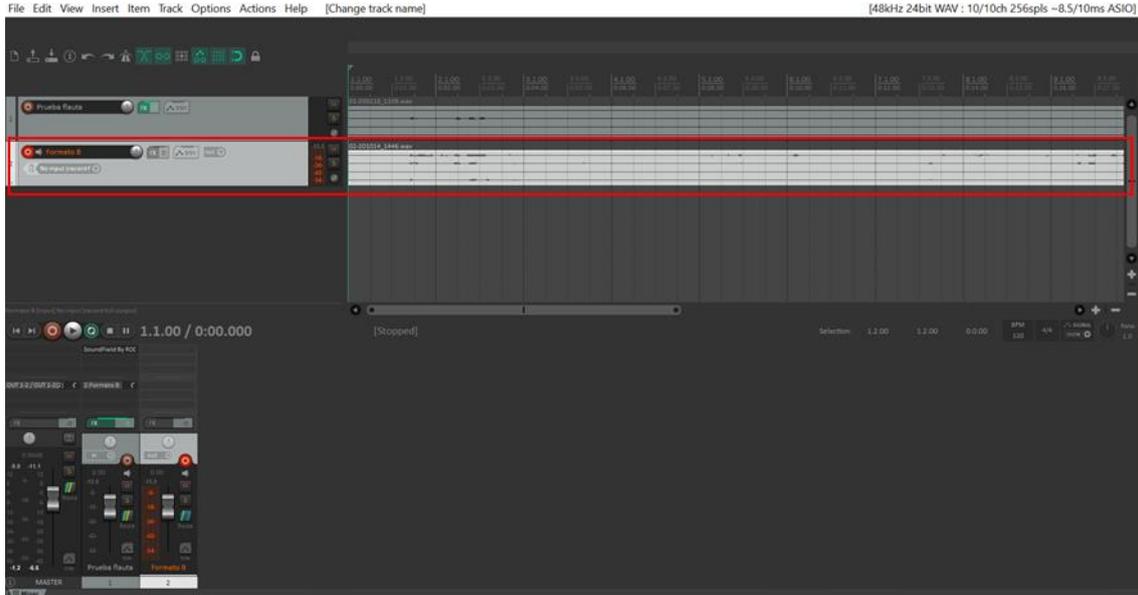


Figura 30. Pista grabada en formato B.

Por último y para poder generar una señal en formato binaural, tenemos que aplicar un filtro de conversión de formatos ya intrínsecos en la aplicación de edición Reaper. El filtro elegido para este proyecto es de sparta. Este filtro ofrece la posibilidad de ajustar los HRIR según se necesite, en este caso utilizaremos los parámetros definidos de fábrica que se muestran en la figura 31.

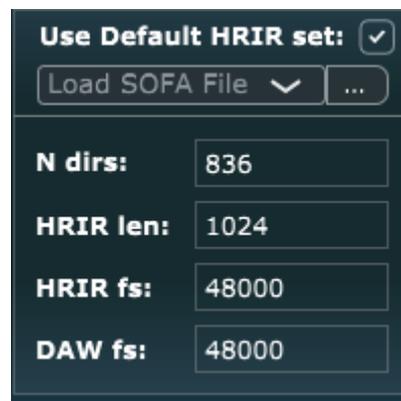


Figura 31. Parámetros conversión a binaural.

La aplicación del filtro es idéntica a la mencionada para el paso del formato A to B, se debe de generar una tercera pista receptora de la salida del nuevo filtro para albergar las señales en formato estéreo “L y R” de la grabación, siendo este el formato final binaural.

Sparta_ambiBIN es un decodificador ambisónico binaural (hasta el séptimo orden) con un cargador SOFA incorporado y soporte de seguimiento de cabeza a través de mensajes OSC. Incluye: opciones de decodificación de mínimos cuadrados (LS), remuestreo espacial (SPR), alineación temporal (TA) y mínimos cuadrados de magnitud (Mag-LS).

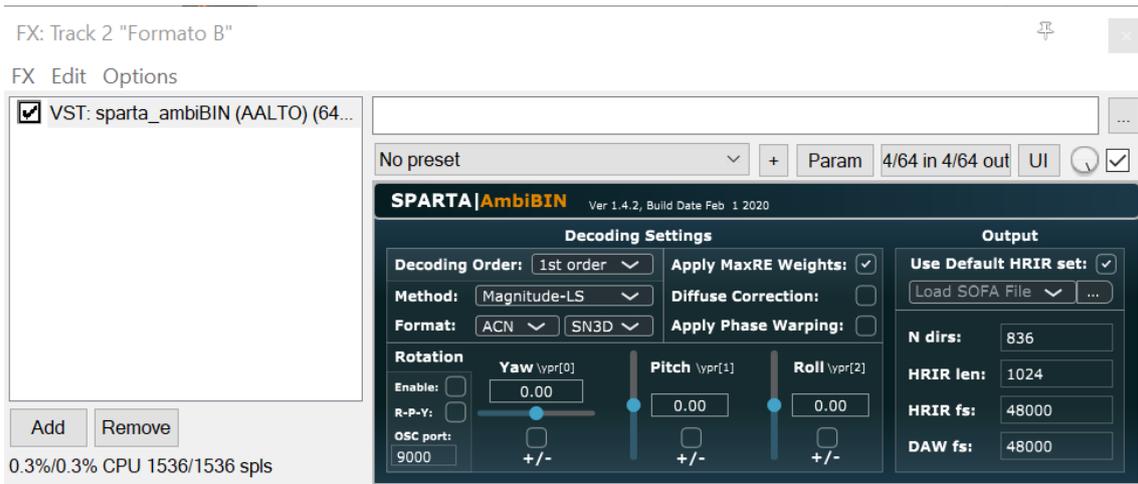


Figura 32. Filtro sparta_ambiBIN

Tras aplicar el filtro y general la ruta hasta la pista número 3 donde se va a albergar la señal binaural de dos canales en formato estéreo, se procede a la grabación.

A continuación, se muestra en la figura 33 la tercera pista con nombre “Binaural”. Esta pista alberga la señal de la pista después de haber pasado por dos filtros anteriormente nombrado y generado en formato estéreo para su posterior reproducción en sistemas estéreo L y R.

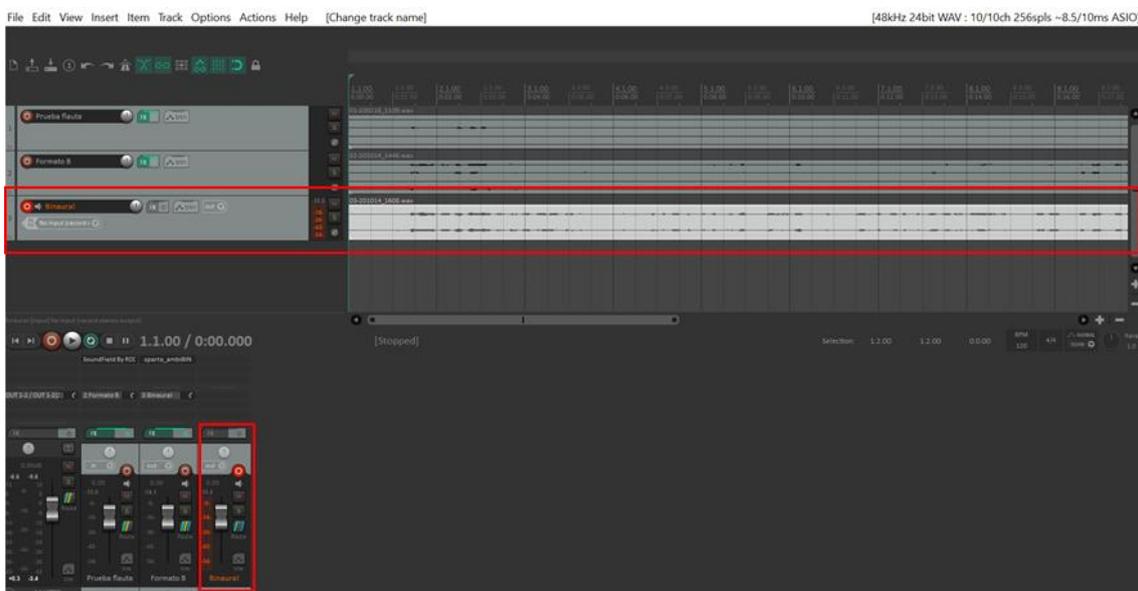


Figura 33. Pista final formato Binaural en estéreo.

Como podemos observar en la Figura 33, la ruta de la segunda pista tiene una entrada y una salida. La entrada referente a salida del primer filtro, el filtro para cambiar a formato B y la salida que se refiere a enviar la salida del filtro ambiBIN a la entrada de la tercera pista.

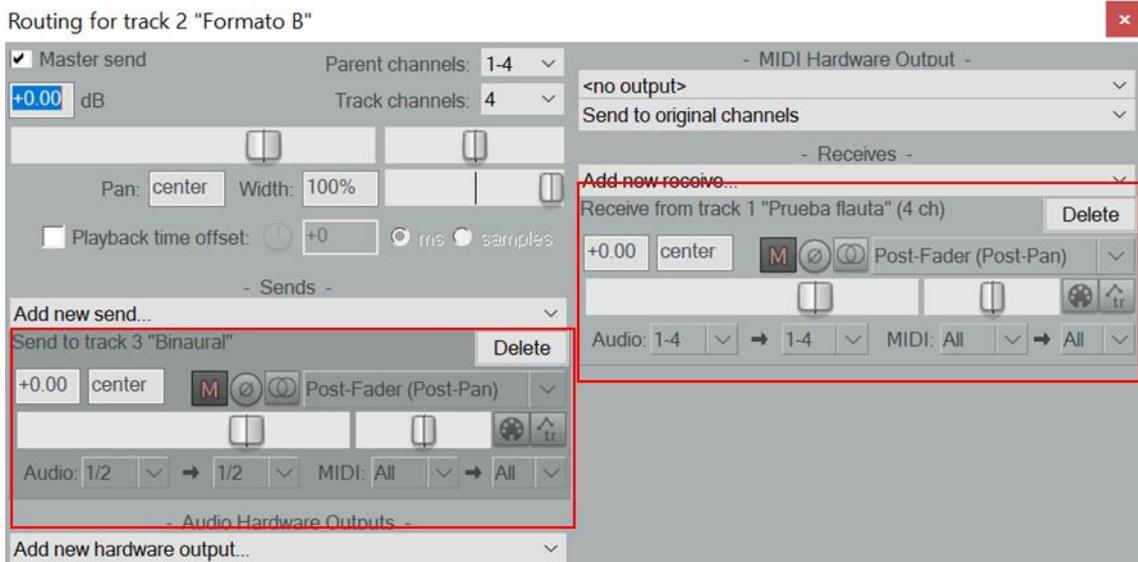


Figura 34. Routing pista 2.

Por último, debemos de renderizar la pista 3 generada con la señal Binaural en formato estéreo. Para ello seguimos las indicaciones del Reaper y generamos la señal en formato .WAV.

A continuación, mostramos paso a paso cómo renderizar en Reaper:

1. Primero abrimos el menú archivo y seleccionamos la opción Renderizar.

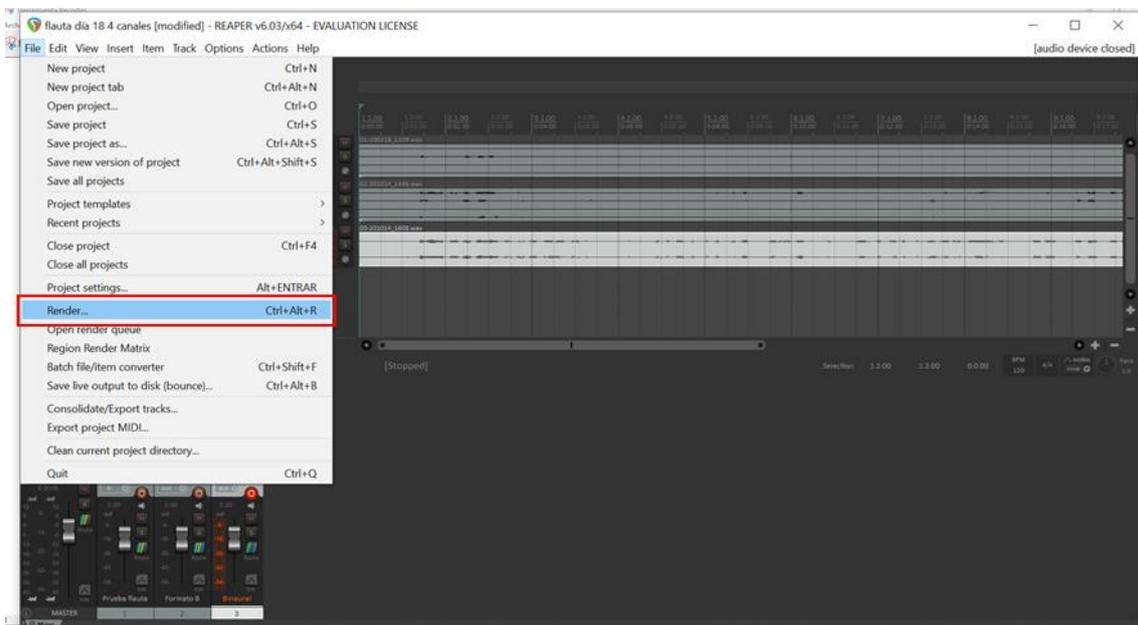


Figura 35. Renderizar 1.

2. A continuación, en la opción Source elegimos la opción común mente conocida en el argot como 'Chorizar' y elegimos la pista que queremos Renderizar, en este caso la número 3.

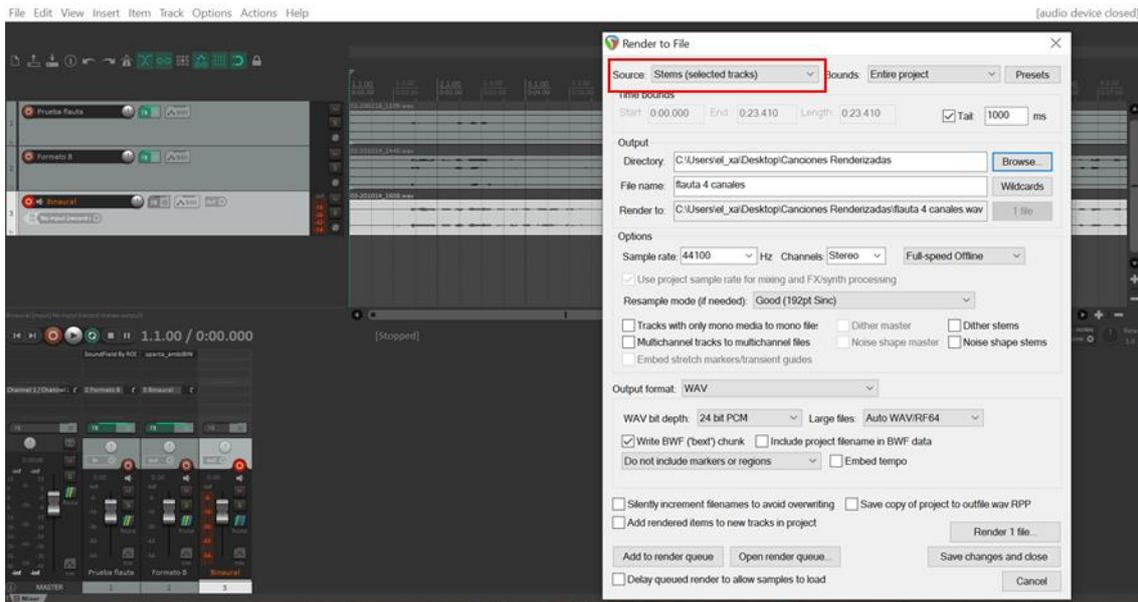


Figura 36. Renderizar 2.

3. Por último, seleccionamos la carpeta donde queremos guardar el archivo, lo nombramos y pulsamos la opción renderizar. Al realizar estas acciones, aparece la pestaña finished con algunos datos sobre el proceso de renderizado que hemos tenido como puede ser el formato del archivo, la frecuencia del sample rate y el tamaño del archivo.

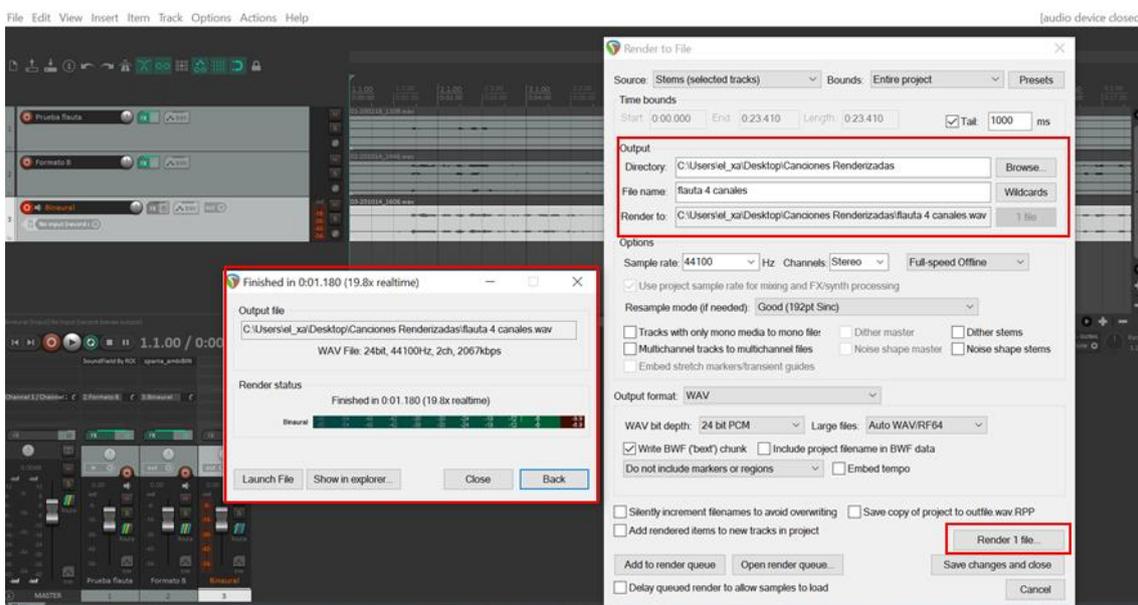


Figura 37. Renderizar 3.

Tras generar el archivo, comprobamos con unos auriculares el sonido binaural que, siendo grabado mediante el uso de 4 micrófonos en tecnología ambisonics, crear para el oyente una sensación de sonido 3D similar a la de estar físicamente en la habitación o el lugar donde se producen los sonidos.

Una vez generado los archivos .WAV de las pruebas en el laboratorio damos por concluido el trabajo en los laboratorios de la Universidad y procedemos con las grabaciones experimentales fuera del laboratorio.

4.1.2 Grabaciones experimentales fuera de laboratorio

Después de las pruebas realizadas en los laboratorios de la universidad se pasó a realizar grabaciones en conjuntos musicales en directo. Debido a la situación a nivel mundial en la que nos encontrábamos, estas grabaciones no pudieron ser lo dinámicas que se esperaban ni tampoco con la variedad que se imaginó al principio de afrontar el presente proyecto.

Dada la situación y el confinamiento, pude contactar con la Agrupación musical “San Clemente de la Mancha” de la localidad en la que resido para solicitar permiso y hacer grabaciones experimentales dentro de sus aulas de ensayo mientras estas varias veces a la semana quedaban para volver a coger el ritmo después de varios meses de inactividad.

La agrupación musical de San Clemente es reconocida a nivel nacional e incluso ha llegado a participar estos años en certámenes internacionales con buenísimos resultados. A continuación, se indican varios de los últimos acontecimientos a nivel nacional que ha tenido la agrupación:

- En abril de 2019 la AM San Clemente de la Mancha viajó a Italia para participar en el concurso internacional "Flicorno D'Or".
- El día 22 de Julio de 2017, la A.M. San Clemente de la Mancha participó en el WMC Kerkrade, en Holanda, a las 18:00 h. en el Parkstad Limburg Theater Heerlen.



Figura 38. Agrupación Musical San Clemente de la Mancha.

Tras hablar con Julián Redondo, director actual de la agrupación musical, se iniciaron las grabaciones experimentales fuera del laboratorio, en este caso en una sala habilitada y reestructurada especialmente para evitar reverberaciones en los ensayos. En la actualidad la banda musical está dividida en varios grupos para realizar las pruebas y ensayos debido a la normativa nacional de reuniones. Para estas grabaciones, se tomaron las medidas de seguridad, distanciamiento e higiene marcadas por los protocolos nacionales.

En la actualidad la agrupación musical está dividida en dos grupos; viento y percusión. Se dividen de esta manera para asegurada las medidas de seguridad, por lo cual es imposible a día de hoy hacer una grabación de la agrupación musical íntegra. Para este proyecto decidí experimentar con la parte de viento ya que acústicamente quedaría mejor en este tipo de grabaciones. La parte de viento consta con una variedad

A continuación, muestro las imágenes de las grabaciones. Los instrumentos que se han grabado, de tipo madera aguda, para los experimentos son los siguientes: flautas, oboes y clarinete bajo.



Figura 39. Madera aguda.

Para realizar las grabaciones se dispuso un sistema de semicírculo en torno al director Julián. Para llegar al objetivo de reproducción binaural, se ha ido moviendo el micro en diferentes posiciones a lo largo de la sala para una vez tratada la señal, tras los filtros correspondientes antes mencionados, notar en nuestros oídos el efecto y saber diferenciar lo cercano o lejano que se encuentra cada tipo de instrumento y la situación en el eje X, Y, Z y Y.

Se han grabado varias canciones del repertorio clásico y barroco del siglo XVIII y XIX. Las canciones tratadas han sido las siguientes:

- Allegro for fluteclock de J. S. BACH.
- Danza de los Mirlitons de Tchaikovsky.
- London trio, de J. Haydn.
- La Muerte de Ase, de Peer Gyn.

En las figuras 40 y 41, se muestran la distribución del personal músico que se ha ofrecido voluntario para realizar estos experimentos y la colocación del material que se ha utilizado para las grabaciones. En este caso, el micro se coloca en frente de los músicos, que están situados en semicírculo mirando al director del grupo musical.



Figura 40. Grabaciones experimentales fuera del laboratorio.



Figura 41. Grabaciones experimentales fuera del laboratorio.

Como se puede comprobar en las figuras 40 y 41 la sala está completamente adaptada con elementos absorbentes que pretenden evitar la reverberación y así evitar el efecto rebote que puede causar estas salas ya que con un buen diseño se puede lograr romper la onda de sonido y con un buen material absorber gran parte del ruido creado.

Tras las primeras pruebas en las cuales se grabaron fragmentos de las canciones mencionadas anteriormente, se decidió a grabar una canción entera que estaba fuera del período de pruebas y experimentos. La canción elegida fue “London trio, de J.Haydn. Algunas de las pruebas realizadas, fragmentos y la canción completa serán añadidas en un anexo final con el conjunto completo de la mayoría de las grabaciones realizadas para la elaboración del presente proyecto.

4.2 Grabaciones a conjuntos acústicos musicales.

Una vez se han probado los instrumentos de grabación en conjuntos musicales, obteniendo resultados satisfactorios en cuanto a grabación, aplicación de filtros y reproducción binaural en auriculares estéreo, se ha dado paso a realizar grabaciones más ‘profesionales’.

Debido a la situación de confinamiento, las restricciones y limitaciones que en esta época nos ha tocado vivir, no se ha podido grabar todo lo que en un principio se pensó, pero sí que se ha logrado contactar con profesionales de la música. En este caso se han ofrecido dos voluntarios profesionales de la música para realizar diversas grabaciones.

4.2.1 *Laura Moreno*

La primera elegida es una cantante que lleva varios años en el sector de la música y ha crecido como profesional de una manera muy rápida. Su nombre es Laura Moreno y es una joven estudiante de la localidad de San Clemente. Con 21 años es cantante y compositora, auto produce sus propios temas, además de tocar la guitarra y el piano de manera autodidacta. Ha realizado conciertos en salas muy conocidas de Madrid como “Buho Real” o “Libertad 8”. A su vez ha sido telonera de grupos conocidos internacionalmente como “Efecto Pasillo”, “Rozalen”, “India Martinez”, etc.

Para esta grabación se ha elegido una única canción la cual se ha grabado en dos distancias para lograr percibir el efecto tras procesar la señal grabada por el micrófono profesional de Rode. La canción elegida se titula “Nunca estoy” del cantante C. Tangana y ha sido acompañada de una guitarra acústica tocada por la propia Laura.

A continuación, se muestran unas imágenes de las grabaciones. La primera grabación se hizo a una distancia de 1 metro de Laura. La segunda grabación se realizó a 2 metros de la cantante. Tras aplicar los filtros correspondientes y poder escuchar el archivo .WAV con unos auriculares comunes, comprobamos la diferencia entre una grabación y otra, distinguiendo sin problemas la lejanía entre una u otra.



Figura 42. Grabación Laura Moreno distancia 1 metro.



Figura 43: Grabación Laura Moreno distancia 2 metros.

Al igual que hemos mostrado en los experimentos, ambas señales grabadas son filtradas y guardadas para posteriormente renderizar y obtener una canción en formato binaural. Tras reproducir ambas canciones podemos distinguir la distancia más o menos lejana de donde se encuentra en este caso la cantante. En las figuras 44 y 45 se muestra un pantallazo del programa utilizado Reaper para el tratamiento de las señales grabadas a Laura.

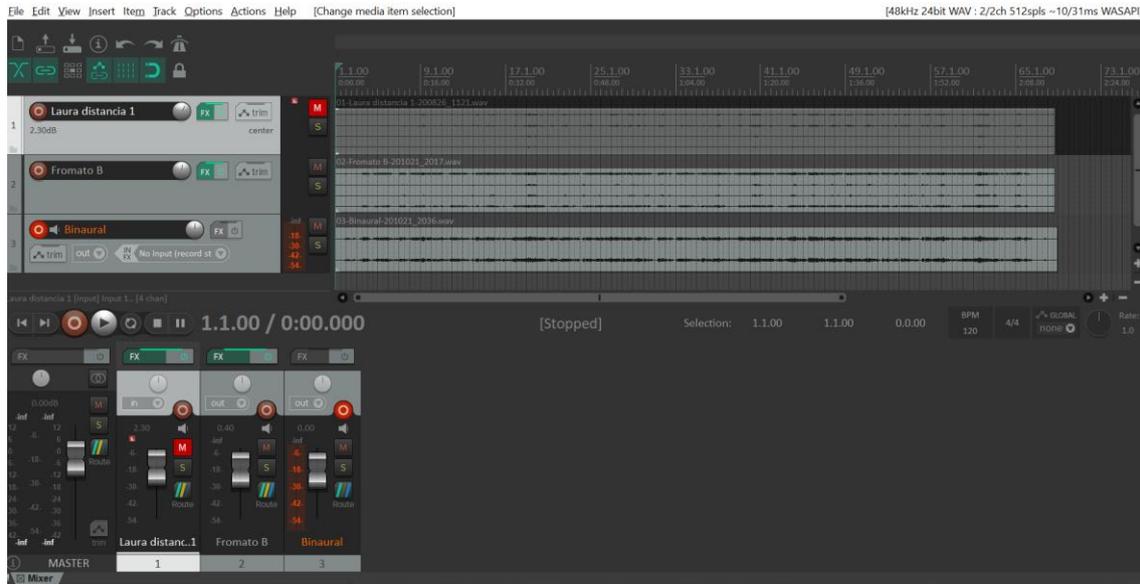


Figura 44. Reaper Laura distancia 1 metro.

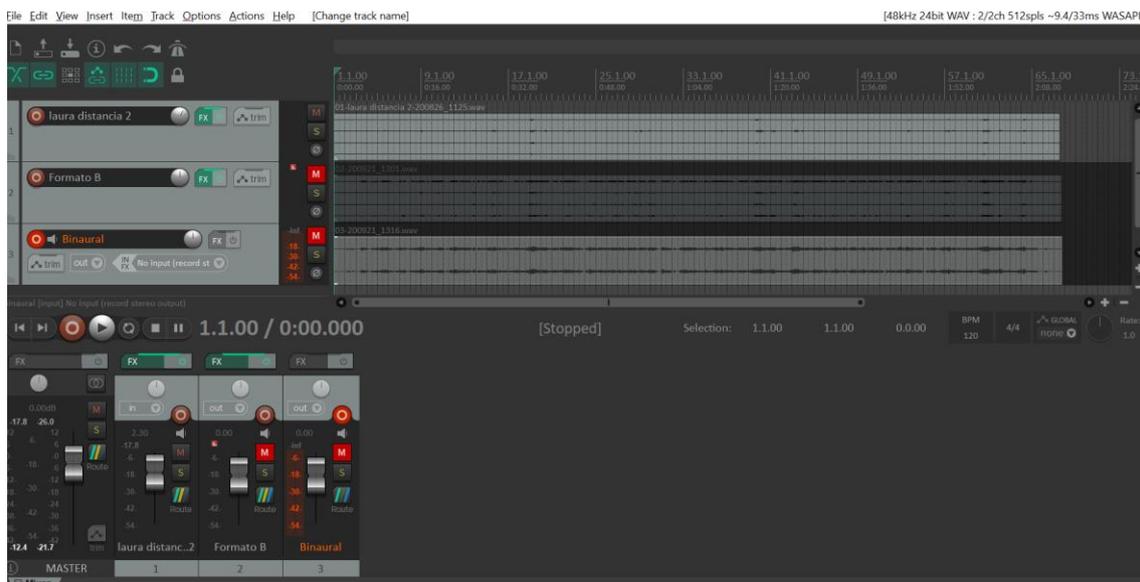


Figura 45. Reaper Laura distancia 2 metros.

Como se puede comprobar observando el nivel de ambas señales, la canción que se graba a una distancia de 1 metro tiene una amplitud mayor en cada una de las pistas guardadas si las comparas con las mismas pistas de la grabación.

4.2.2 Julián Redondo

Al final de realizar las grabaciones experimentales con la banda, con el resultado de una canción entera válida para un conjunto acústico musical, el director de la banda de música de San Clemente se ofreció voluntario para que grabásemos varias canciones.

Julián cuenta con un conocimiento y experiencia extraordinarios. Hijo del antiguo director de la agrupación musical de San Clemente, es graduado profesional como aspirante a matrícula de honor en su especialidad. En 2017 finaliza sus enseñanzas superiores en el conservatorio superior de música de Castilla La Mancha con sede en Albacete. Durante toda su carrera desde el año 2008 como músico ha participado en innumerables concursos, colaboraciones y actuaciones en el ámbito nacional que le han llevado a ser un gran profesional de la música.

Para las grabaciones en conjuntos musicales se han elegido tres instrumentos de percusión a los cuales el músico tiene gran aprecio y los conoce muy bien: Marimba, Lira y Vibráfono que se muestran a continuación.



Figura 46. Marimba



Figura 47. Lira



Figura 48. Vibráfono

Esta vez se decide hacer grabaciones tanto en cercanía y lejanía como también en diferentes alturas para estudiar mejor el efecto de la grabación con tecnología ambisonic en conjuntos musicales.

4.2.2.1 Marimba

Este instrumento ha sido el más utilizado en las grabaciones debido a su gran versatilidad, se le conoce como un xilófono de láminas de madera, siendo un instrumento de percusión auditivo con las láminas dispuestas de mayor a menor y cada una con una altura de sonido diferente para ser golpeadas con mazos. Las características principales de la Marimba son las siguientes:

- Extensión de LA (1) a Do(6). 4 1/3 octavas.
- Afinación: La = 442Hz.
- Soporte graduable en altura desde 85 a 100 cm.
- 4 Ruedas direccionables con freno.
- Láminas de madera de padouk de 64-40mm.
- Tubos de resonancia de aluminio.
- Medidas: 200 cm. de largo x 88 cm. de ancho.
- Nivel Profesional.
- Totalmente desmontable.

Las canciones elegidas y tocadas por Julian son las siguientes:

- Charles Chaplin
- Fuga en la menor de J.S. Bach
- Giga de J.S. Bach
- Porgy and Bess
- Danza de los Mirlitons de Tchaikovsky

Estas canciones han sido grabadas en distintas distancias al igual que se ha hecho para el resto de las grabaciones para una vez tratadas en Render y reproducidas en nuestros oídos mediante auriculares estéreo comprobar y saber distinguir las diferencias entre una y otra en cuanto a distancia, posición y altura.

A continuación, se muestran en las siguientes figuras el momento de Julián tocando este instrumento para las grabaciones realizadas.



Figura 49. Julián tocando Marimba grabación de cerca.



Figura 50. Grabación Julián Marimba micro alto.

4.2.2.2 Lira

Elegimos la Lira ya que es un instrumento casi mítico, aunque no muchos conocen para qué ha sido utilizado. En este caso, la Lira fue el instrumento elegido para tocar la canción mundialmente conocida de Harry Potter. Este instrumento es muy peculiar como así demuestra su sonido al ser percutido. Estas liras se han convertido en imprescindibles para las mejores bandas y orquestas, con un timbre brillante, lleno de armónicos que se proyectan a través de cualquier ensemble. Este instrumento tiene las siguientes características:

- Láminas de 31mm fabricadas en acero de alta calidad con perfil Round Top y niquelado satinado.
- Octavas (C5-E8).

- El sistema de pedal reversible permite configurar el apagado.
- Tubos resonadores de acero niquelado brillante y montados en soportes de goma aislante.
- Mueble de madera natural lacada.
- Bastidor Traveler, robusto y ajustable en altura entre 80cm y 100cm.
- Incluye cubierta para las láminas.
- Afinada a 442Hz.

A continuación, se muestran imágenes de Julián percutiendo la Lira para tocar la canción de Harry Potter.



Figura 51. Grabación Julián Lira.

4.2.2.3 Vibráfono

Por último, se hicieron las grabaciones con el instrumento Vibráfono, otro instrumento de percusión. Este instrumento es similar al xilófono, pero con la diferencia de que las láminas están fabricadas en aluminio. Las láminas son percutidas con baquetas o mazos donde su cabeza está fabricada con hilo, lana o nailon. Este instrumento esconde un motor para que el sonido se alargue mucho más, y además provoca ondas por las vueltas que da el rodillo de debajo de las láminas que crea un efecto diferente al que hace el vibráfono sin el motor, que son ondas más planas. Las características principales del vibráfono son las siguientes:

- Motor: 40-140 rpm
- Rango: 3.1 octavas E3-F6
- Láminas: Aluminio 50-38 mm
- Afinación: 442 Hz.
- Marco: Voyager
- Longitud: 156 cm
- Anchura máxima: 75 cm
- Anchura mínima: 40 cm
- Ajuste de altura: 85 a 105 cm

A continuación, se muestran las imágenes de las grabaciones de Julián percutiendo en el Vibráfono la canción “Chega de Saudade” del compositor T. Jobim.



Figura 52. Grabación Julián Vibráfono.

Capítulo 5. Conclusiones

En este proyecto se ha explorado el funcionamiento del sistema ambisonic, un sistema relativamente novedoso y complejo para la grabación del sistema espacial.

El alumno ha adquirido conocimientos del funcionamiento de la técnica del proceso tanto a nivel matemático como a nivel práctico. A su vez se ha utilizado hardware de última generación para realizar las grabaciones de sonidos utilizando un micrófono ambisonic que ha necesitado ser configurado. Se han instalado los plugins correspondientes en el software adecuado y se han superado las dificultades que han ido surgiendo hasta manejar con soltura todo tipo de software y hardware necesario para realizar los experimentos con amplio conocimiento de causa.

Como conclusiones del trabajo, se ha obtenido un know-how muy importante sobre estas técnicas de grabación al mismo tiempo que se han obtenido una serie de resultados y grabaciones que pueden ser muy útiles para el futuro estudio del funcionamiento del sistema. Por otro lado, también se han estudiado las técnicas de conversión de ambisonic a formato binaural, técnicas muy populares hoy en día para el sonido 3D. Para ello se han utilizado unos plugins que convierten la señal WXYZ a una señal HRTF que permite ser oída con auriculares.

Con todo esto se puede concluir de manera satisfactoria el desarrollo del presente proyecto obteniendo resultados muy interesantes en cuanto a la grabación y procesado con tecnología ambisónica.

Capítulo 6. Bibliografía

- [1] Mario Ledo Hernansanz “Tecnología de registro y reproducción sonora para entornos de realidad virtual” Proyecto Fin de Carrera,UPM, Jun. 2013
- [2] <https://victorperales.wordpress.com/2015/11/02/ambisonics-como-alternativa/>
- [3] Moreau, S., Daniel, J., & Nicol, R. (2003). Further Investigations Of High Order Ambisonics And Wavefield Synthesis For Holophonic Sound Imaging. AES 114th Convention Amsterdam.
- [4] D.G.Malham “Spatial hearing mechanisms and sound reproduction” Music technology group, University of York, England, 1998
- [5] A. R. Moller, Hearing: Anatomy, physiology, and disorders of the auditory system, Plural Publishing Inc., 2012.
- [6] C. I. Cheng y G. H. Wakefield, «Introduction to head related transfer functions (HRTFs): Representations of HRTFs in Time, Frequency and Space,» de AES 107th convention, New York, 1999.
- [7] J. Blauert, «Sound Localization in the Median Plane,» Acta acustica, vol. 22, nº 4, 1969.
- [8] D. H. Mershon y L. E. King, «Intensity and reverberation as factors in the auditory perception of egocentric distance,» Perception & Psychophysics, vol. 18, nº 6, pp. 409-415, 1975
- [9] A. J. Kolarik, B. C. J. Moore, P. Zahorik, S. Cirstea y S. Pardhan, «Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss,» Atten Percept Psychophys, vol. 78, pp. 373-395, 2016.
- [10] Hugeng, D. Gunawan y W. Wahab, «Effective Preprocessing in Modeling Head-Related Impulse Responses Based on Principal Components Analysis,» Signal Processing: An International Journal (SPIJ), vol. 4, nº 4, pp. 201-212, 2010
- [11] H. Lehnert y J. Blauert, «Principles of binaural room simulation,» Applied Acoustics, vol. 36, nº 3-4, pp. 259-291, 1992.
- [12] <https://www.waves.com/ambisonics-explained-guide-for-sound-engineers>
- [13] <http://www.rode.com/>

Capítulo 7. Anexos

7.1 Anexo A: Archivos Reaper en formato .rpp

- Danza de los Marlitons (Banda)
- La muerte de Ase de Peer Gyn
- Laura Moreno
- Lira Harry Potter
- London trio de J. Hydn (Banda)
- Marimba Charles Chaplin
- Marimba Fuga en la menor de J. S. Bach
- Marimba Giga de J. S. Bach
- Marimba Porgy and Bess
- Prueba Flauta (Laboratorio)
- Prueba Javi Guitalla (Laboratorio)
- Vibráfono Chega de Saudade de T. Jobim

7.2 Anexo B: Canciones Binaural

- Danza de los Marlitons
- La muerte de Ase de Peer Gyn
- Laura Moreno
- Lira Harry Potter
- London Trio de J. Hydn
- Marimba Charles Chaplin
- Marimba Fuga en la menor de J. S. Bach
- Marimba Giga de J. S. Bach
- Marimba Porgy and Bess
- Prueba Flauta
- Prueba Javi Guitarra
- Vibráfono Chega de Saudade de T. Jobim

7.3 Anexo C: Fotos y Vídeos

- Fotos Banda
- Fotos Juli
- Fotos Laura
- Vídeos Juli y Laura