



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

TELECOM ESCUELA
TÉCNICA **VLC** SUPERIOR
DE INGENIERÍA DE
TELECOMUNICACIÓN

Estudio y optimización de técnicas de grabación de sonido 3D mediante Ambisonics

Fernando Ayelo Sánchez

Tutor: José Javier López Monfort

Trabajo Fin de Máster presentado en la Escuela Técnica Superior de Ingeniería de Telecomunicación de la Universitat Politècnica de València, para la obtención del Título de Máster Universitario en Ingeniería de Telecomunicación

Curso 2020-21

Valencia, 23 de junio de 2021

Escuela Técnica Superior de Ingeniería de Telecomunicación
Universitat Politècnica de València
Edificio 4D. Camino de Vera, s/n, 46022 Valencia
Tel. +34 96 387 71 90, ext. 77190
www.etsit.upv.es

**VLC/
CAMPUS**
VALENCIA, INTERNATIONAL
CAMPUS OF EXCELLENCE



Resumen

Con la llegada de nuevas tecnologías en alza en el mundo multimedia como la realidad virtual o la realidad aumentada, y de los avances dentro del audio para cine y ficción, ha sido necesario poder ofrecer un sistema de sonido espacial preciso y efectivo con un flujo de trabajo adaptado a los tiempos que corren. Por ello, y con más de 40 años de historia y en plena expansión de sus capacidades, la tecnología Ambisonics se vuelve un firme candidato para cubrir estas necesidades dentro del mundo del sonido espacial y ofrecer una solución abierta y modular, que además engloba los procesos de grabación, almacenamiento, mezcla y reproducción de audio espacial.

Por todo ello, en el presente proyecto se va a realizar un estudio completo de la tecnología de sonido espacial Ambisonics. Se hará un repaso a la escucha humana y como esta es capaz de identificar los sonidos y sus direcciones así como a los fundamentos del sonido espacial y sus características. Además se hará un breve repaso a otras tecnologías de sonido espacial como el *surround*, Dolby®ATMOS o Auro3D.

A continuación se realizará un estudio exhaustivo sobre Ambisonics y su teoría, los fundamentos sobre los que se sostiene, herramientas disponibles para trabajar con esta tecnología, tanto hardware como software, flujo de trabajo y características que la definen.

Se verán los distintos formatos de reproducción y grabación en Ambisonics, así como posibles aplicaciones para dichos procesos o cosas a tener en cuenta en caso de que queramos grabar o reproducir material Ambisonics de una forma profesional y optimizada.

Por último se realizarán una serie de mezclas en formato Ambisonics para mostrar como trabajar con esta tecnología, aspectos importantes a tener en cuenta durante el proceso y herramientas disponible y su uso. Para evaluar estas mezclas y el proceso llevado a cabo se buscará destacar el potencial de dicha tecnología para tareas de mezcla, especialmente musical y evaluar estos aspectos y los resultados obtenidos.

Palabras clave: Audio, Ambisonics, Atmos, Sonido Espacial, Campo Sonoro, Binaural, Reaper

Resum

Amb l'arribada de noves tecnologies en alça en el món multimèdia com la realitat virtual o la realitat augmentada, i dels avanços dins de l'àudio per a cinema i ficció, ha sigut necessari poder oferir un sistema de so espacial precís i efectiu amb un flux de treball adaptat als temps que corren. Per això, i amb més de 40 anys d'història i en plena expansió de les seues capacitats, la tecnologia Ambisonics es torna un ferm candidat per a cobrir aquestes necessitats dins del món del so espacial i oferir una solució oberta i modular, que a més engloba els processos d'enregistrament, emmagatzematge, mescla i reproducció d'àudio espacial.

Per tot això, en el present projecte es realitzarà un estudi complet de la tecnologia de so espacial Ambisonics. Es farà un repàs a l'escolta humana i com aquesta és capaç d'identificar els sons i les seues adreces així com als fonaments del so espacial i les seues característiques. A més es farà un breu repàs a altres tecnologies de so espacial com el *surround*, Dolby®ATMOS o Auro3D.

A continuació es realitzarà un estudi exhaustiu sobre Ambisonics i la seua teoria, els fonaments sobre els quals se sosté, eines disponibles per a treballar amb aquesta tecnologia, tant hardware com software, flux de treball i característiques que la defineixen.

Es veuran els diferents formats de reproducció i enregistrament en Ambisonics, així com possibles aplicacions per a aquests processos o coses a tindre en compte en cas que vulguem gravar o reproduir material Ambisonics d'una forma professional i optimitzada.

Finalment es realitzaran una sèrie de mescles en format Ambisonics per a mostrar com treballar amb aquesta tecnologia, aspectes importants a tindre en compte durant el procés i eines disponible i el seu ús. Per a avaluar aquestes mescles i el procés dut a terme es buscarà destacar el potencial d'aquesta tecnologia per a tasques de mescla, especialment musical i avaluar aquests aspectes i els resultats obtinguts.

Paraules clau: Audio, Ambisonics, Atmos, So Espacial, Camp Sonor, Binaural, Reaper

Abstract

With the arrival of new technologies on the rise in the multimedia world such as virtual reality or augmented reality, and the advances in audio for film and fiction, it has become necessary to be able to offer a precise and effective spatial sound system with a workflow adapted to this times. Therefore, and with more than 40 years of history and in full expansion of its capabilities, Ambisonics technology becomes a strong candidate to meet these needs in the world of spatial sound and offer an open and modular solution, which also encompasses the processes of recording, storage, mixing and playback of spatial audio.

Therefore, in this project, a complete study of Ambisonics spatial sound technology will be carried out. It will look at human listening and how it is able to identify sounds and their directions as well as the fundamentals of spatial sound and its characteristics. In addition, a brief overview of other spatial sound technologies such as surround, Dolby®ATMOS or Auro3D will be given.

This will be followed by an exhaustive study of Ambisonics and its theory, the foundations on which it is based, the tools available to work with this technology, both hardware and software, workflow and its defining characteristics.

The different Ambisonics recording and playback formats will be discussed, as well as possible applications for these processes or things to consider in case we want to record or playback Ambisonics material in a professional and optimised way.

Finally, a series of Ambisonics mixes will be performed to show how to work with this technology, important aspects to take into account during the process and available tools and their use. In order to evaluate these mixes and the process carried out, the aim will be to highlight the potential of this technology for mixing tasks, especially musical, and to evaluate these aspects and the results obtained.

Keywords: Audio, Ambisonics, Atmos, Spatial Sound, Sound Field, Binaural, Reaper

Índice general

1	Introducción	1
1.1	Motivación	1
1.2	Objetivos	2
2	Fundamentos del sonido espacial	3
2.1	Psicoacústica: el oído humano y la localización espacial	3
2.1.1	El oído: partes y funcionamiento	3
2.1.2	Localización espacial	5
2.2	Audio multicanal: Los inicios del audio espacial	14
2.2.1	Estéreo	15
2.2.2	Cuadrafónico	17
2.2.3	Dolby y los sistemas multicanal	17
3	Ambisonics: Teoría y Fundamentos	20
3.1	Grabación. Micrófonos Ambisonics.	21
3.1.1	Micrófonos HOA	23
3.2	Orden de canales y normalización.	24
3.2.1	Orden de los canales	24
3.2.2	Normalización	25
3.3	Conversión A/B. Fuentes Ambisonics.	27
3.4	Codificación. Fuentes mono y estéreo.	30
3.5	Formato B. Upmix, postproducción y mezcla.	31
3.5.1	Aplicación de efectos no lineales en HOA mediante SPS	32
3.6	Decodificación y <i>playback</i>	37
3.6.1	Binaural	37
3.6.2	<i>Playback</i> en arreglos de altavoces	40
3.7	Distribución Ambisonics.	46

3.7.1	Formatos de archivos Ambisonics. Metadatos.	48
4	Aplicación del Ambisonics en Grabación y Reproducción	51
4.1	Reproducción	51
4.1.1	5.1 binaural	52
4.1.2	<i>Head-tracker</i> DIY	56
4.1.3	Laboratorio de pruebas	58
4.2	Grabación	66
4.2.1	Situación del micrófono	67
4.2.2	Interfaz propietaria	69
4.2.3	Interfaz de audio	71
4.2.4	PC (Software)	72
5	Mezcla en Ambisonics	78
5.1	Caso práctico 1: Ambisonics + <i>spot</i>	79
5.2	Caso práctico 2: Multipista acústico	83
5.2.1	Caso práctico 2.1 - Proyecto Ambisonics	84
5.2.2	Caso práctico 2.2 - Proyecto estéreo	87
5.3	Caso práctico 3: Multipista sintético	88
6	Resultados	90
6.1	Resultados caso práctico 1: Ambisonics + <i>spot</i>	90
6.2	Resultados caso práctico 2: Multipista acústico	91
6.2.1	Resultados Caso práctico 2.1 - Proyecto Ambisonics	92
6.2.2	Resultados Caso práctico 2.2 - Proyecto estéreo	92
6.2.3	Resultados generales Caso práctico 2	93
6.3	Resultados caso práctico 3: Multipista sintético	94
7	Conclusiones y trabajos futuros	96
7.1	Conclusiones	96
7.2	Propuestas de trabajo futuras	97
	Bibliografía	97
	Anexos	101
A	Fórmulas para armónicos de orden superior (HOA).[19]	102

B	Coordenadas esféricas del Laboratorio de Percepción Espacial del Sonido	103
C	Sistemas de reproducción Ambisonics en el Laboratorio	105
C.1	Sistemas Dolby Surround	105
C.2	Sistemas Ambisonics	106

Índice de figuras

1.1	Logotipo Ambisonics.[1]	1
2.1	Fisiología del oído humano.[2]	4
2.2	Diferencia interaural de tiempo.	6
2.3	Planos relativos al oyente y ángulos de posicionamiento de la fuente.[4]	7
2.4	ITD “dentro” y “fuera” del mismo ciclo de onda.	8
2.5	Efecto sombra o <i>head shadow effect</i> .	8
2.6	Cono de confusión. Las fuentes 1-2 y 3-4 poseen los mismos tiempos ITD y los mismos niveles ILD para el oyente.	10
2.7	PRTF.	11
2.8	Efecto <i>Doppler</i> .	13
2.9	Fonógrafo de la marca Edison.[6]	14
2.10	Cilindro de cera negra de la marca Edison de 1904.[7]	14
2.11	Gramófono con bocina de amplificación.[8]	15
2.12	Disco de vinilo de 12” a 33 RPM.[9]	15
2.13	Diagrama del prototipo del Théâtrophone de 1881.[10]	16
2.14	Ilustración de un Théâtrophone de <i>Le Magasin Pittoresque</i> (1892).[10]	16
2.15	Principio de la técnica de captación estereofónica Blumlein.[11]	16
2.16	Ejemplo de sistema Dolby Surround 7.1 para salas de cine.[16]	18
3.1	Armónicos esféricos hasta orden $n = 4$.[17]	20
3.2	Flujo de trabajo dentro de Ambisonics.	22
3.3	Sistema de coordenadas antropométrico según norma ISO 2631.	22
3.4	Micrófonos Ambisonics HOA. De izda. a dcha. <i>Eigenmike® em32</i> , <i>CoreSound OctoMic</i> y <i>Eigenmike® em64</i> .	23
3.5	Respuesta teórica del filtro F_W en amplitud y fase para una separación $r=15mm$ entre cápsulas.[18]	28
3.6	Respuesta teórica del filtro F_{XYZ} en amplitud y fase para una separación $r=15mm$ entre cápsulas.[18]	28

3.7	Micrófonos Ambisonics de 1 orden. De izda. a dcha. <i>CoreSound Tetramic, Rode NT-SF1, Sennheiser Ambeo y Soundfield SPS200</i>	29
3.8	VST del Rode NT-SF1 para conversión de formato A-B.	30
3.9	Box externo del Soundfield ST450 MKII para conversión y monitorización de formato A-B.	30
3.10	<i>StereoEncoder</i> del Institute of Electronic Music and Acoustics.	30
3.11	<i>AmbiENC</i> de la suite SPARTA, desarrollado por el laboratorio acústico de la AALTO University.	30
3.12	<i>Upmixer</i> de la suite SPARTA con una configuración de primer a tercer orden.	32
3.13	<i>EnergyVisualizer</i> de la suite IEM para tercer orden.	32
3.14	<i>Eigenmike em32</i> y la ordenación de sus cápsulas similar al SPS-32.[38]	33
3.15	Patrón polar cardioide para diferentes órdenes.[38]	33
3.16	<i>B-Format</i> de 2º orden procedente de <i>OctoMic</i> de <i>CoreSound</i>	34
3.17	<i>AmbiDEC</i> configurado para convertir de 2º orden a <i>T-design</i> de 12 canales.	35
3.18	SPS en formato <i>t-design</i> de 12 canales.	35
3.19	Perfil de ruido obtenido al final de uno de los canales de la grabación.	36
3.20	<i>AmbiENC</i> configurado para convertir de T-12 a Ambisonics 2º orden.	36
3.21	Resultado final procesado con 9 pistas <i>B-format</i>	36
3.22	Sistema de captación binaural <i>Neumann KU-100</i>	38
3.23	Sistema de captación binaural <i>3DIO FreeSpace</i>	38
3.24	Descripción del fenómeno del <i>Crosstalk</i>	39
3.25	Menú de carga para HRTF en el reproductor <i>VLC</i>	39
3.26	Decodificador binaural de Google: <i>Resonance Audio monitor</i>	39
3.27	<i>AmbiDEC</i> de la suite SPARTA con decodificación <i>AllRAD</i> para ambas bandas de trabajo.	44
3.28	<i>Izda</i> . Altavoz en posición nadir imaginario. <i>Dcha</i> . Altavoz en posición nadir real.	46
3.29	Arreglo hemisférico 8+4 en 3D.	46
3.30	Estructura de archivo para los formatos AmbiX básico y extendido. [31]	50
4.1	Ruteo de señal para virtualizar un sistema 5.1.	52
4.2	Interface virtual de audio 5.1.	53
4.3	Configuración del orden de canales del sistema 5.1.	53
4.4	Interface virtual de audio para monitorización de la salida de Reaper.	54
4.5	Configuración interna de VLC para la reproducción en 5.1.	54
4.6	Bus de Reaper para la captura del audio procedente de VLC.	55
4.7	Bus de Reaper para el procesado de audio en Ambisonics.	55
4.8	<i>AmbiENC</i> , de la suite Sparta Ambisonics. Codificador Ambisonics.	56

4.9	AmbiBIN, de la suite Sparta Ambisonics. Decodificador binaural Ambisonics.	56
4.10	Esquema de conexionado entre la placa Arduino Pro Micro y la placa sensora.	57
4.11	SceneRotator de la suite IEM para la rotación del campo sonoro Ambisonics.	58
4.12	Modelo 3D del laboratorio de pruebas levantado en <i>SketchUp</i>	59
4.13	Laboratorio de "Percepción Espacial del Sonido".	59
4.14	Altavoces <i>M-Audio-BX5 D2</i> . https://m-audio.com/products/view/bx5-d2	60
4.15	MOTU 24I/O.	61
4.16	MOTU PCI-424.	62
4.17	Vista frontal del rack con las 4 tarjetas MOTU 24I/O.	63
4.18	Vista trasera del rack con las 4 tarjetas MOTU 24I/O y el cableado necesario.	63
4.19	Diagrama de conexionado del laboratorio.	64
4.20	Excel para decodificar arreglos al laboratorio.	65
4.21	Flujo habitual de grabaciones Ambisonics.	66
4.22	<i>Rode NT-SF1 con suspensión y anti viento de serie</i> . ^[40]	67
4.23	<i>TSL Soundfield ST450 mkii</i> montado sobre cesta anti viento <i>Cinela Pianissimo</i> . ^[39]	67
4.24	<i>Rode NT-SF1</i>	68
4.25	<i>TSL Soundfield ST450 mkii</i> y su <i>box</i> de control.	68
4.26	Posibles posiciones para un micrófono Ambisonics. ^[41]	68
4.27	Interfaz propietaria con conversión AB y monitorización del <i>ST450 mkii</i>	69
4.28	Interfaz propietaria EMIB. (<i>Arriba.- Frontal/Centro.-Trasero/Abajo.-Trasero con interfaz MADI</i>).	71
4.29	Flujo de señal en grabaciones Ambisonics FOA.	72
4.30	VST del <i>Rode NT-SF1</i> . para manejo y conversión entre formatos. Incluye decodificación.	73
4.31	VST <i>Surround Zone 2</i> de <i>TSL Soundfield</i> para manejo de sus micrófonos analógicos.	73
4.32	Interfaz del VST <i>em32setGain</i> . ^[42]	74
4.33	Ventana con la calibración de las cápsulas del <i>em32</i> . ^[42]	74
4.34	Posible diagrama de bloques para trabajar con el <i>em32</i> y las herramientas <i>EigenUnits</i> ®. ^[43]	74
4.35	<i>EigenUnits</i> ® en formato VST en el DAW <i>Reaper</i> . ^[43]	75
4.36	Ventana principal del software independiente <i>EigenStudio</i> ®. ^[44]	76
5.1	Diagrama de situación de los micrófonos con los tiempos de <i>delay</i> necesarios y la lista de canales correspondiente.	80
5.2	Mezclador con los <i>plugins</i> insertados, los colores del diagrama y el bus <i>Ambisonics</i> general.	81
5.3	<i>Room Encoder</i> del paquete IEM para codificación de una fuente Ambisonics dentro una sala.	82
5.4	Vista general con todos los <i>plugins</i> aplicados en el bus AMBISONICS.	82

5.5	Ruteo de la señal enter los diferentes buses y lista de canales.	84
5.6	Diagrama con la ubicación espacial de cada instrumento y lista de canales.	86
5.7	Diagrama con la ubicación espacial de cada instrumento y lista de canales a lo largo del espacio estéreo.	87
5.8	Ventana sesión del caso práctico 3 con todas las envolventes y su automatización. . . .	88
6.1	<i>IEM EnergyVisualizer</i> durante una parte de la mezcla del caso práctico 1.	91
6.2	<i>IEM EnergyVisualizer</i> para mezcla estéreo.	93
6.3	<i>IEM EnergyVisualizer</i> para mezcla estéreo.	93
6.4	<i>IEM EnergyVisualizer</i> para orden 3.	94
6.5	<i>IEM EnergyVisualizer</i> para orden 5.	94
C.1	Dolby 5.0.	105
C.2	Dolby 7.0.	105
C.3	Dolby 9.0.	105
C.4	Dolby 11.0.	105
C.5	1 ^{er} orden (90°).	106
C.6	2 ^o orden (60°).	106
C.7	3 ^{er} orden (45°).	106
C.8	4 ^o orden (40°).	106
C.9	5 ^o orden (30°).	106
C.10	6 ^o orden (20°).	106
C.11	7 ^o orden (15°).	106

Capítulo 1

Introducción

1.1 Motivación

Desde sus comienzos y con el paso de la evolución, el ser humano ha ido desarrollando una escucha binaural con la que no solo pudiese escuchar el entorno que le rodea, sino además poder situar los sonidos dentro del espacio sonoro.

En los principios de las grabaciones sonoras únicamente éramos capaces de realizar grabaciones monofónicas, que si bien representaban el sonido grabado de una forma correcta, distaban mucho de una escucha natural y cercana a la realidad binaural sobre la que se asienta la escucha humana. A estas grabaciones monofónicas le siguieron técnicas de grabación y mezcla en formato estéreo gracias al gran *Alan Blumlein*, creador de la estereofonía, primera técnica de grabación que intentaba acercarse al sonido espacial. El estéreo recreaba una espacialidad limitada que no llegaba a representar un campo sonoro de 360° , si bien se convirtió en lo que hoy todavía conocemos, un estándar de escucha para mezcla y reproducción de material sonoro. Paralelamente al desarrollo y crecimiento del estéreo, en la década de los psicodélicos años 70, fue cuando apareció una tecnología que prometía ofrecer al fin un sonido espacial real y cercano a la escucha humana. Esta tecnología se denominó **Ambisonics**.

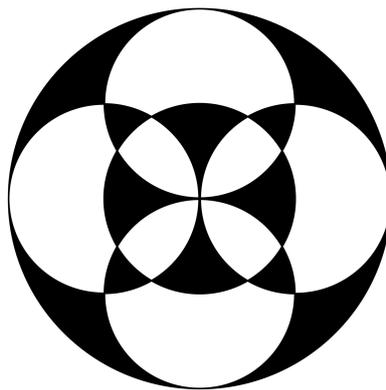


Figura 1.1: Logotipo Ambisonics.[1]

Nacida en la década de los 70 en Reino Unido, con *Michael Gerzon*, del Instituto Matemático de Oxford; el *Dr. Peter Barnes Fellgett*, del Departamento de Cibernética de la Universidad de Reading y *John Stuart Wright* de IMF Electronics como padres, la tecnología denominada Ambisonics no es nueva. Permaneció casi olvidada a lo largo de 40 años, siendo investigada y utilizada únicamente por un pequeño grupo de entusiastas y audiófilos. Es ahora, durante la última década, cuando todo su

potencial ha explotado. Sería lógico preguntarse cuál ha sido el motivo de esta explosión, de este resurgimiento de una tecnología que, si miramos hacia atrás, siempre ha sido relegada al cajón de unos pocos interesados en seguir investigándola.

Los motivos son varios, y variados. Se podría decir que el primer responsable de esta nueva era ha sido la explosión de la realidad virtual (a partir de ahora VR) y de la realidad aumentada (RA). Empresas punteras como Google, Facebook o Youtube (filial de Google) han abanderado la tecnología Ambisonics como la elegida para la reproducción de contenido en 360° en sus plataformas multimedia.

Por otro lado, los avances en materia de procesamiento de señal y el avance en CPU, FPGA¹ o los DSP² también tienen parte de culpa. Esto es debido a que la tecnología Ambisonics hace gran uso de este tipo de procesamiento para ser efectiva, así como para conseguir unos resultados óptimos y realistas tanto en las grabaciones, como en la reproducción de contenido con audio 3D.

1.2 Objetivos

El presente proyecto presenta varios frentes u objetivos que buscan ser alcanzados a la finalización del mismo.

El primer objetivo consiste en realizar un repaso al actual estado de una tecnología como Ambisonics de cara a poder afrontar el resto de objetivos con una base teórica y práctica lo suficiente amplia como para entender ciertos aspectos que la rodean. Entre esos aspectos destaca la localización espacial que realiza el cerebro humano de los sonidos que le rodean, o la teoría de los armónicos esféricos sobre la que se sustenta Ambisonics.

Una vez entendida y repasada la tecnología Ambisonics, se busca diseñar y mostrar diferentes sistemas de reproducción como la virtualización de sistemas 5.1 en formato Ambisonics, el desarrollo de un pequeño *head-tracker* basado en Arduino o el diseño e implementación de diferentes sistemas de reproducción 3D en el laboratorio de pruebas del departamento GTAC de la UPV. Además, se busca ofrecer una serie de consejos y técnicas para la grabación de material a partir de microfónica Ambisonics, especialmente la disponible en la UPV.

Por otro lado, y debido a las limitaciones para realizar grabaciones de campo debido a la situación del COVID-19, se verá cómo trabajar con material Ambisonics en la tarea de mezcla y edición del mismo. Para ello, se busca poder realizar distintas mezclas a partir de material sonoro variado en formato Ambisonics y ver qué resultados somos capaces de obtener.

El objetivo final del presente proyecto es el de al acabar, comprender de manera general y profunda cómo funciona la tecnología Ambisonics y cómo podemos trabajar con ella de una manera profesional, atendiendo en todo momento a sus virtudes y sus defectos para obtener de ella el mayor potencial posible.

¹FPGA.- Field-Programmable Gate Array.

²DSP.- Digital Signal Processor.

Capítulo 2

Fundamentos del sonido espacial

El marco teórico se va a dividir en dos grandes bloques temáticos para una mayor comprensión y organización de la información.

El primer gran bloque va a versar sobre psicoacústica, siempre desde el punto de vista que nos interesa en el presente proyecto, la localización espacial. Cómo el ser humano, por sus condiciones fisiológicas y evolutivas, posee un sistema de localización de los sonidos en el espacio con unas características y unas limitaciones que a menudo vamos a tener que contemplar y tener presentes mientras trabajamos con audio espacial

Por otro lado se presentará un segundo gran bloque centrado única y exclusivamente en audio multicanal. Cómo los sistemas de audio más primitivos fueron evolucionando desde la estereofonía propuesta por *Alan Blumlein* en la década de los 30 hasta tecnologías actuales y sofisticadas como Dolby Atmos o Auro-3D.

2.1 Psicoacústica: el oído humano y la localización espacial

Antes de ver en qué consiste el audio espacial, y en especial Ambisonics, se va a hacer un repaso a los mecanismos que tiene y utiliza el cerebro para localizar y ubicar los sonidos que le llegan dentro del plano sonoro. Se hablará de la Teoría Dúplex para recordar y entender que trucos y herramientas utiliza el cerebro de manera inconsciente para ubicar los sonidos que le llegan en uno u otro sitio dentro del plano sonoro.

2.1.1 El oído: partes y funcionamiento

Si bien es cierto que este proyecto versa sobre audio espacial, y por tanto, la parte que más nos interesa es cómo nuestro oído es capaz de localizar los sonidos en el espacio, es importante ver en primer lugar de qué partes está formado y cuáles son las tareas principales de cada una de esas partes. Para ello, nuestros oídos, actúan a modo de “sensores”, encargados de recoger el sonido que les llega y manipularlo para que pueda ser entendido por el cerebro. Esto es un proceso que se ha ido mejorando de manera evolutiva a lo largo de nuestra historia, debido en parte a nuestro rol dentro de la naturaleza.

Para ello, el oído cuenta con tres partes fundamentales y bien diferenciadas: el oído externo, medio e interno.

Oído externo

Formado por dos elementos principales bien diferenciados: en primer lugar se encuentra el pabellón auricular, pabellón auditivo o pinna, comúnmente conocido como oreja y gracias a su forma helicoidal llena de pliegues y canales, es un elemento clave a la hora de recoger y encaminar los frentes de onda que le llegan hacia el interior, donde se encuentra el canal auditivo externo.

Éste, consiste en un conducto o canal, de entre 25 y 30 mm de longitud y 7 mm de diámetro, y el cuál presenta una resonancia alrededor de los 3 kHz. Dicha resonancia tiene sentido de manera evolutiva si pensamos en qué banda de frecuencias se encuentra la mayor cantidad de información del habla humana. Gracias a dicha resonancia, se mejora la inteligibilidad de la palabra de una manera pasiva, todo gracias a la forma física de nuestro oído externo, el cuál será clave para la localización espacial del sonido, como ya se verá más adelante.[2]

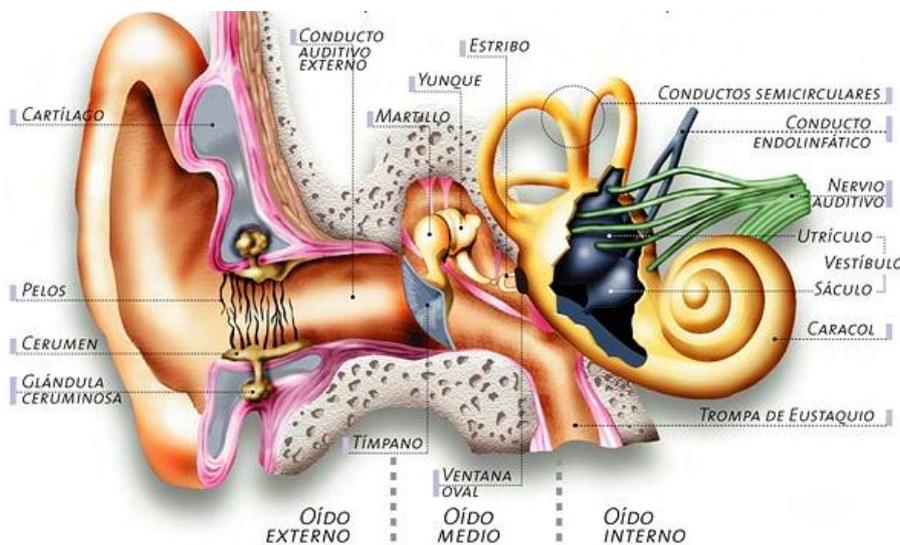


Figura 2.1: Fisiología del oído humano.[2]

Oído medio

Una vez el sonido atraviesa el canal auditivo, éste incide contra el tímpano, separador natural entre el oído externo y medio. El tímpano consiste en una membrana, similar a la membrana de un micrófono, que vibra cuándo el sonido le llega, actuando por tanto como un transductor acústico-mecánico.

Una vez el tímpano recoge los frentes de onda y vibra, transmite estas vibraciones a una cadena de huesecillos(también llamados osículos auditivos). Dicha cadena de huesecillos está formada por el martillo, el yunque y el estribo, el hueso más pequeño del cuerpo humano. Esta cadena de huesos de pequeño tamaño actúan como un compresor de audio, de manera que las vibraciones con mucha amplitud se atenúan, protegiendo así al resto del oído de fuertes impactos sonoros. Es por esto, que cuándo estamos en un ambiente muy ruidoso, y pasamos a uno muy silencioso, tardamos un tiempo en acostumbrarnos, ya que la cadena de huesecillos tiende a relajarse.

Una vez atravesados los osículos, las vibraciones se transmiten hasta la ventana oval, también conocida como la puerta al oído interno.

Otro de los elementos presentes en el oído medio es la trompa de Eustaquio, encargada de conectar el oído medio con la rinofaringe para igualar la presión a ambos lados del tímpano.[2]

Oído interno

El oído interno empieza donde termina el medio, en la ventana oval. Junto con la cóclea o caracol, debido a su forma externa, ambos forman el oído interno.

La cóclea a su vez, y en su interior, está formada por una serie de membranas encargadas del reconocimiento frecuencial de los sonidos que nos llegan, trabajo que realiza la membrana basilar y el órgano de Corti.

El funcionamiento es el siguiente: las vibraciones comprimidas a través de la cadena de huesecillos es transmitida sobre la ventana oval a través de vibraciones, es decir, el sonido ya se ha convertido en vibraciones mecánicas. Una vez llegado a la ventana oval, éstos se transmiten al interior de la cóclea, donde el líquido linfático que la llena comienza a moverse por las vibraciones. Dichas vibraciones a lo largo de la membrana basilar, la cuál es capaz de “descomponer” el sonido a lo largo de su espectro frecuencial, activan el órgano de Corti, el cuál se encarga de realizar la conversión de las vibraciones en impulsos nerviosos para ser enviados al cerebro a través del nervio auditivo, actuando así de transductor mecánico-eléctrico.

La membrana basilar consiste en una especie de alfombra compuesta por unas 25000 células pilosas, las cuáles son excitadas gracias a la vibración del líquido linfático. Dicha membrana es más ligera y rígida cerca de la ventana oval, y más pesada y elástica según nos alejamos de ella, lo cuál permite que la frecuencia de resonancia dentro del órgano de Corti vaya variando. De esta manera, a medida que nos alejamos de la ventana oval, la membrana basilar resuena a frecuencias más graves, cubriendo así todo el espectro audible.

Por tanto, dependiendo de la amplitud y de la composición espectral de un sonido entrante, se estimularán más o menos células pilosas a lo largo de la membrana basilar. Éstas células codificarán sus estímulos a través del órgano de Corti en estímulos nerviosos, los cuáles se envían al cerebro a través del nervio auditivo. De esta manera tan compleja, el oído humano es capaz de escuchar e interpretar hasta 10 octavas, y con una sensibilidad de entre 0 dB (umbral de audición) y 130 dB (umbral del dolor).[2][3]

Tras este breve y conciso resumen sobre el oído humano y su fisionomía, ya somos capaces de entender como el ser humano es capaz de transformar los sonidos que le llegan en información válida e interpretable por el cerebro, pero todavía no somos capaces de saber como localiza dichos sonidos en el espacio.

2.1.2 Localización espacial

Antes de comenzar a hablar de cómo el cerebro interpreta las señales que le llegan para poder determinar la localización espacial de un sonido dentro del campo sonoro, es importante hablar de los dos grandes tipos de escucha que poseemos y las diferencias entre ambas.

Por un lado, tenemos la escucha monoaural, es decir, la utilización de un sólo oído. Es lógico pensar que no es un tipo de escucha habitual, ya que los seres humanos, por norma general poseemos dos oídos. Se habla sin embargo de escucha monoaural como forma de aclarar el concepto para futuras menciones.

Por otro lado tenemos la escucha binaural, es decir, la que se da por la interacción de ambos oídos. Por tanto, no se trata únicamente de la información que manda un oído al cerebro, si no de la que mandan ambos y las diferencias entre ambas, las que ofrecen información al cerebro relevante a la posición de la fuente en el espacio.

Indicios interaurales

Una vez vistos los dos tipos de escucha posibles, llegamos al primer pilar de la localización espacial: las diferencias interaurales. Éstas funcionan como mecanismos de localización binaural, debido a que es el cerebro quien compara la información que le llega de ambos oídos y es mediante las diferencias entre éstas que extrae la información necesaria para localizar la fuente dentro del espacio. Uno de los puntos a destacar de este tipo de localización espacial es que sólo es capaz de localizar fuentes dentro del plano azimutal u horizontal, de manera que si una fuente se encuentra fuera del plano horizontal que rodea al oyente, éste no será capaz de determinar su posición con precisión

Fue la **Teoría Dúplex**, iniciada por *Lord Rayleigh* (1877) con su libro “The Theory of Sound” la primera en sugerir que los principales mecanismos de localización espacial utilizados de manera inconsciente por el cerebro provienen de la escucha binaural, concretamente de las diferencias interaurales de tiempo (Interaural Time Differences o ITD) y de las diferencias interaurales de nivel (Interaural Level Differences o ILD).[2]

ITD: Las diferencias interaurales de tiempo es uno de los principales mecanismos del cerebro a la hora de determinar la posición de una fuente dentro del espacio. Estudios llevados a cabo por *Wightman y Kistler* (1992) mostraron cómo la ITD puede llegar a ser una de las pistas de localización espacial más sólidas entre las conocidas.

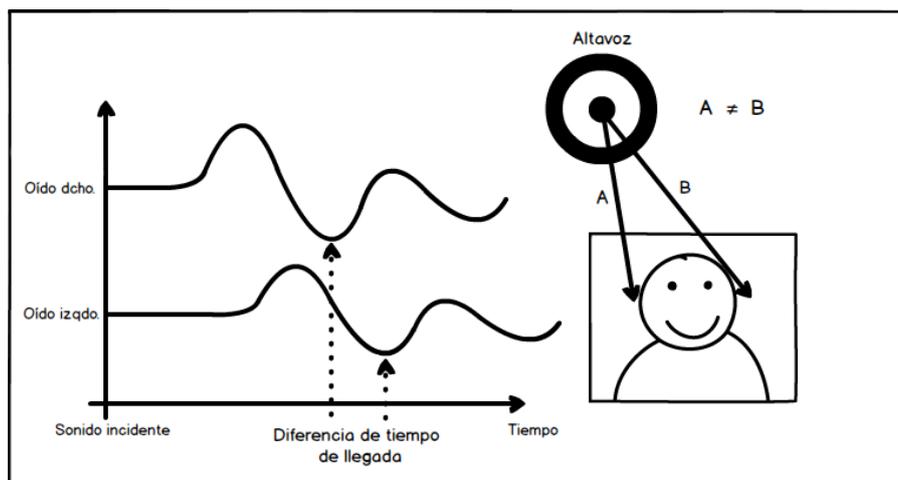


Figura 2.2: Diferencia interaural de tiempo.

Tal y como se observa en la figura 2.2, una fuente situada fuera del eje central del sujeto, es decir, si tomásemos como punto de referencia su punto frontal, siempre que la fuente estuviera desplazada de dicho eje, existirá una diferencia de tiempo desde que el sonido llega a un oído, hasta que llega al otro. Es el cerebro quien interpreta esa diferencia para determinar la posición de la fuente.

Para situar los elementos dentro del plano, se toma como referencia el sistema de planos mostrado en la figura 2.3. Se utilizarán coordenadas esféricas, definiendo la posición de la fuente respecto al oyente a través de una distancia (r) y dos ángulos, el primero de ellos medido sobre el plano horizontal, ángulo de azimut (θ), tomando el frente como 0° . El otro ángulo se eleva sobre el plano horizontal y mide la pendiente del vector con su origen en el centro de la cabeza del oyente, ángulo de elevación (φ).

Es por tanto, que gracias a la diferencia de tiempos, y por tanto, de fase, que el cerebro es capaz de determinar la posición de una fuente. Por todo ello las ITD pueden variar desde un mínimo de 0

segundos de diferencia cuando la fuente sonora posee un azimut de 0° , hasta aproximadamente 0.69 segundos para una fuente con un ángulo azimut de 90° , es decir, colocada en un lateral del oyente.

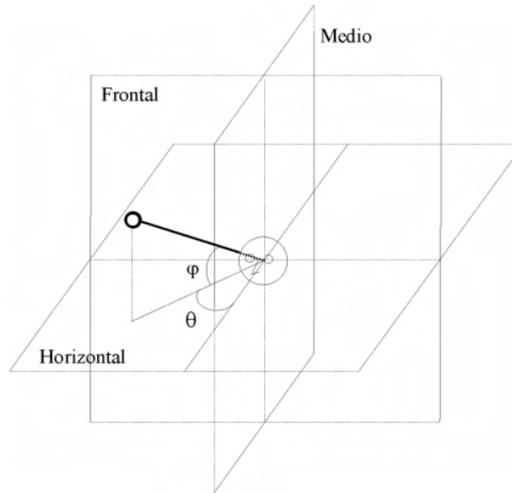


Figura 2.3: Planos relativos al oyente y ángulos de posicionamiento de la fuente.[4]

Ahora bien, hasta ahora hemos hablado de tiempo, pero no de frecuencia. La longitud de onda varía con la frecuencia, de modo que para frecuencias graves, con una longitud de onda grande, ambos oídos se encontrarán escuchando la señal dentro del mismo ciclo. El problema viene cuando la frecuencia aumenta, ya que la longitud de onda disminuye y por tanto a cada oído llegará un ciclo de onda distinto, provocando en el oyente incertidumbre y confusión.

Tomando el diámetro medio de una cabeza humana, que está aproximadamente en torno a los 23 cm para un ser humano adulto, podemos determinar en que frecuencia está el límite que determina cuándo escucharemos dentro del mismo ciclo, y cuándo la longitud de onda será tan pequeña que no superará esa distancia.

$$\lambda = 23\text{cm} = 0,23\text{m} \text{ diámetro medio de una cabeza humana}$$

$$\lambda = \frac{v}{f} \rightarrow f = \frac{v}{\lambda}$$

Donde $v = 331 * (0,6 * T)$ siendo T la temperatura ambiente

Por tanto, para una temperatura media de 20° tendremos una velocidad del sonido de 343 m/s

Por tanto nos queda que:

$$f = \frac{v}{\lambda} = \frac{343}{0,23} \simeq 1500\text{Hz}$$

Si bien es cierto que en el punto en el que la longitud de onda corresponde a aproximadamente el doble del diámetro de la cabeza, es decir, cuándo la longitud de onda es de unos 0.46 m, 750 Hz aproximadamente, ya existen problemas de localización, cuando se alcanzan los 1500 Hz es cuando realmente el indicador ITD se vuelve prácticamente inútil, ya que la información que proporciona al cerebro se torna confusa e inservible para la localización de la fuente.

Este problema viene derivado de que el cerebro no es capaz de discernir qué frente de ondas le ha llegado primero al oído, pudiendo incluso llegar a cada oído un frente de ondas diferente pero con la misma fase, tal y como se puede observar en la figura 2.4, por lo que llegados a este límite de los 1500 Hz, la única manera que tiene el cerebro de distinguir diferencias entre ambos oídos es a través de la

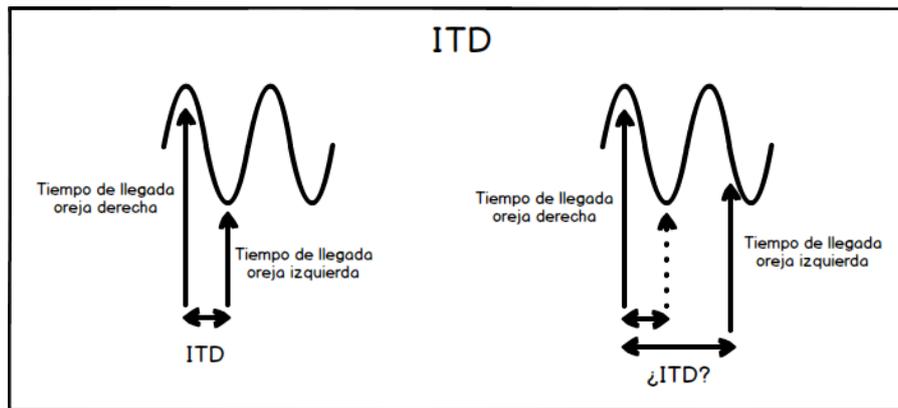


Figura 2.4: ITD “dentro” y “fuera” del mismo ciclo de onda.

diferencia de amplitud que llega a cada oído, apareciendo en este punto otro indicador utilizado para la localización espacial para frecuencias superiores a los 1500 Hz mencionados.

ILD: Las diferencias interaurales de nivel dan al cerebro la información necesaria para localizar una fuente sonora en el espacio en frecuencias por encima de los 1500 Hz de límite mencionados en el apartado 2.1.2. Esto quiere decir que, mientras que por debajo del límite frecuencial mencionado, la información para la localización espacial se extrae del ITD, es por encima de dicho límite cuando el ILD nos proporciona la información que el ITD no es capaz de darnos por los problemas comentados anteriormente.

Su funcionamiento es lógico e intuitivo, y consiste en la comparación de energía que nos llega a cada uno de los dos oídos de manera involuntaria por el cerebro. Otro de los factores que proporcionan información al cerebro es el *head shadow effect* o **efecto sombra**.

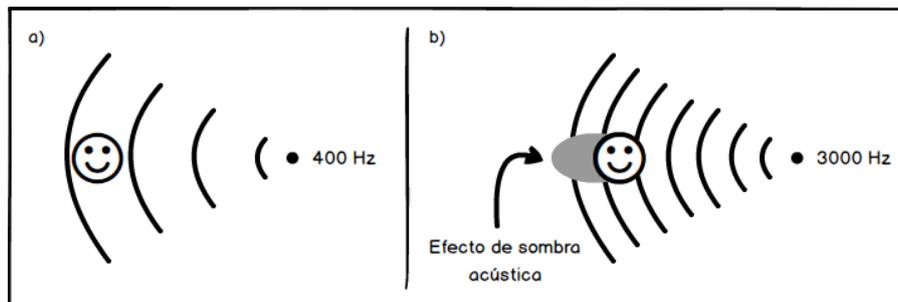


Figura 2.5: Efecto sombra o *head shadow effect*.

El efecto sombra se da siempre a partir de los 1500 Hz mencionados y aumenta con la frecuencia, siendo más pronunciado a medida que ésta aumenta. Para ondas con una longitud mayor a la cabeza, cuyo límite viene marcado en 1500 Hz, el efecto es irrelevante, siendo la cabeza un obstáculo cuasi transparente. Para las ondas con una longitud igual o inferior al tamaño de la cabeza, ésta se convierte en un obstáculo insalvable, siendo más relevante a medida que la frecuencia aumenta. Este efecto viene explicado por dos leyes fundamentales de la acústica, la ley de **Masa Acústica**, y la difracción del sonido.

- **Ley de Masa Acústica:** la ley nos dice que en una barrera acústica, como es la cabeza, a mayor masa de ésta, mayor será la absorción del sonido y por tanto menor la transmisión, siendo dicha reducción de unos 6 dB por octava para una cabeza humana media. [2]

- **Difracción del sonido:** Uno de los grandes fenómenos que afectan al sonido es la difracción. El sonido posee la capacidad intrínseca de poder rodear un obstáculo, donde a menor frecuencia, mayor difracción y viceversa. Por tanto, un sonido de baja frecuencia será capaz de rodear un obstáculo menor que su longitud de onda sin verse afectado, mientras que a frecuencias altas, el sonido no podrá sortear dicho obstáculo, de donde se producirán reflexiones, refracciones y absorciones, alterando así la trayectoria del mismo, y pudiendo incluso llegar a generar nuevos frentes de onda a partir de la propia difracción.

Por tanto, y en vista de lo comentado, podemos resumir que para frecuencias superiores a los 1500 Hz, el ILD aporta al cerebro el *grosso* de información necesaria para una correcta localización espacial. Por contra, para frecuencias graves éste se vuelve inútil, y es el ITD quien aporta la información en esta zona del espectro, complementándose y cubriendo todo el espectro entre ambos.

Efecto Haas: También conocido como **efecto de precedencia** o ley del primer frente de ondas, el efecto Haas, denominado así por el médico alemán *Helmut Haas* (1951), quien describió el efecto de precedencia por primera vez, consiste en un efecto binaural psicoacústico protagonista de gran importancia en la localización espacial por parte del cerebro. Es en la situación en la que, ya sea por un eco o un rebote por ejemplo, un sonido es seguido de otro, si la diferencia de tiempo en la que ambos sonidos llegan al oído es suficientemente corta, el cerebro tomará como fuente del sonido únicamente el primer sonido, siendo el segundo totalmente desechado. Se puede tomar por tanto como un fenómeno claro de enmascaramiento psicoacústico por parte del cerebro, volviéndose imposible de diferenciar un sonido de otro y tomando como fuente y referencia únicamente el primero.

Dicho enmascaramiento no es fijo para un tiempo concreto, ya que su apreciación vendrá relacionada con la envolvente del sonido, teniendo una expresión máxima entre 0.6 y 10 ms. Por ejemplo, para sonido con unos transitorios muy marcados, el efecto se hará notable hasta los 15 o 20 ms, tiempo a partir del cuál se harán apreciables las dos fuentes. Sin embargo, si se trata de sonidos con transitorios lentos, es decir, con un tiempo de ataque y decaimiento más largos, el tiempo de diferencia entre fuentes para que éstas sean totalmente identificables aumenta hasta los 50 o 70 ms.

Por tanto, el efecto Haas nos aporta una información notable a la hora de situar fuentes sonoras en el espacio, y será más o menos apreciable dependiendo tanto del intervalo de tiempo de diferencia entre las fuentes, como de la envolvente del sonido producido por éstas.

Para poder recrear este efecto y apreciar sus consecuencias sobre el sonido y la posición virtual que el cerebro interpreta de las fuentes, se puede aplicar, sobre una pareja de monitores correctamente ajustados, un retardo de entre 5 y 50 ms en uno de los altavoces. Se apreciará como para un retardo de 0 ms, la fuente virtual se sitúa en el centro, mientras que a medida que aumentamos el retardo, ésta se va desplazando en primer lugar hacia el monitor que no tiene el retardo aplicado, es decir, la fuente que llega primero. Llegará un punto en que el retardo sea tan grande que en vez de detectar una sola fuente llegando de un punto desplazado, empezaremos a escuchar dos fuentes claramente diferenciadas a cada lado del monitor. Será éste el intervalo de tiempo en el que el efecto Haas dejará de surtir efecto, y por tanto, el tiempo límite en el que el cerebro dejará de interpretar ambas fuentes como una sola y empezará a apreciarlas como dos fuentes separadas. [2]

Cono de confusión: Como ya se ha comentado a lo largo de este apartado, los parámetros ITD e ILD aportan al cerebro información sobre la localización espacial dentro del plano azimutal u horizontal, pero todavía no se ha explicado como localizar fuentes dentro del plano de elevación, debido en parte a que dichos parámetros poseen ciertas limitaciones.

Una de ellas es el denominado **cono de confusión**. Tal y como se aprecia en la figura 2.6, las

fuentes 1-2 y 3-4 poseen un mismo valor de ILD e ITD, haciendo imposible al cerebro discernir de dónde procede el sonido, creando así una sección de espacio sonoro con forma de cono en donde es imposible identificar la posición exacta de la fuente. Esta zona o región dentro del espacio se conoce como cono de confusión debido a la forma que tiene, es una de las principales limitaciones de los parámetros ILD e ITD para la localización espacial.

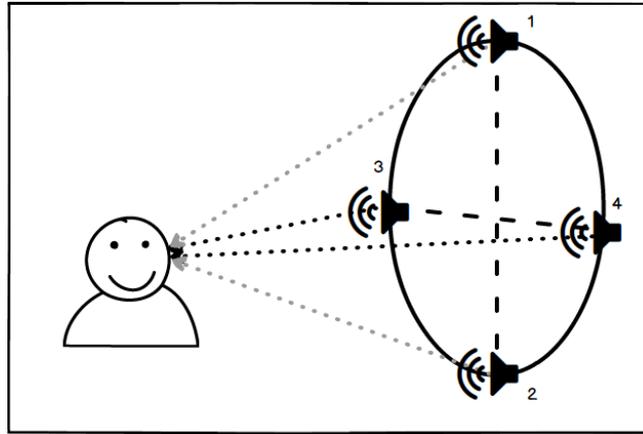


Figura 2.6: Cono de confusión. Las fuentes 1-2 y 3-4 poseen los mismos tiempos ITD y los mismos niveles ILD para el oyente.

Indicios monoaurales

Tal y como se ha visto dentro del apartado 2.1.2, existen una serie de indicadores interaurales para la localización espacial dentro del plano azimutal u horizontal, que son los encargados de aportar al cerebro la información necesaria para identificar la posición de una fuente dentro del plano sonoro. Estos indicadores tienen limitaciones, en especial dentro de la localización fuera del plano horizontal, es decir, con fuentes elevadas o por debajo del plano horizontal de escucha. Para poder solucionar estas confusiones o incógnitas, el cuerpo humano hace uso de otro tipo de indicadores, en este caso monoaurales, para ser capaz de detectar la posición espacial de una fuente.

HRTF: *Head-Related Transfer Function*. El oído, órgano encargado de la audición, actúa como un filtro con una respuesta concreta y variable para cada persona. Este filtro es utilizado en muchas ocasiones por el cerebro humano para resolver posibles conflictos con la posición de una fuente sonora dentro del cono de confusión, ayudando al cerebro a situar la fuente en el espacio.

Para ello, el cerebro utiliza éste indicio monoaural para, a partir de las diferentes reflexiones y difracciones provocadas tanto por la oreja o *pinna*, como por la cabeza entera o el torso, determinar la posición de la fuente. Es por tanto que la HRTF se comporta como un filtro único y personal a través del cuál se modifica el espectro frecuencial de la señal original, provocando retardos y cambios de fase para algunas frecuencias concretas, generándose nuevos frentes de onda y provocando cancelaciones si alguno de esos nuevos frentes de onda llega en contra fase con el sonido directo.

Tal y como se observa en la figura 2.7 (se ha representado únicamente la oreja para el ejemplo, PRTF¹, pero la HRTF incluye toda la cabeza y el torso), dependiendo de la dirección de llegada del sonido, éste sufrirá unas alteraciones concretas, pudiendo provocar cancelaciones o sumas para según

¹PRTF.- *Pinna-rated Transfer Function*.

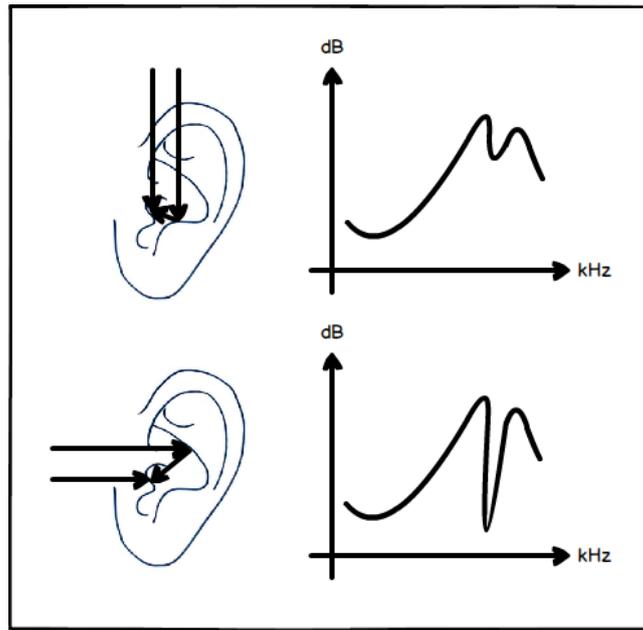


Figura 2.7: PRTF.

que frecuencias, alterando así el espectro del sonido original y ayudando al cerebro a identificar la posición de la fuente para posiciones elevadas entre otras.

Por tanto, cuando trabajamos con auriculares para mezcla de audio espacial, todas esas reflexiones y difracciones provocadas por nuestro cuerpo no se dan, perdiendo así información que necesita el cerebro para la ubicación de las fuentes. Para solucionar esta ausencia de información, se utiliza el uso de filtros HRTF medidos en laboratorio para simular todos esos fenómenos acústicos. Como es lógico, lo ideal es utilizar una HRTF personal y propia medida en laboratorio, que contenga el comportamiento acústico natural del sujeto que la utilice, de manera que su cerebro interprete la información a la salida del filtro de manera correcta. Como esto no siempre es posible por la dificultad de la medición, existen respuestas HRTF genéricas para su uso en cualquier software que permita cargar el archivo *SOFA*, formato utilizado para el uso de HRTF en software de audio.

Ahora bien, el uso de éstas HRTF genéricas no siempre es lo idóneo, ya que aparecen problemas graves a la hora de localizar sonidos fuera de la cabeza (internalización), así como confusiones frente-atrás y viceversa (inversión). Esto es debido a que nuestro cerebro, después de toda una vida escuchando, “conoce” nuestra HRTF, y la utiliza para la localización, volviéndose confusa si esa HRTF es otra. Estas limitaciones son importantes a la hora de mezclar o trabajar con audio 360° con HRTF genéricas.[5]

Para la obtención de la HTRF, esta se puede hallar a través de la transformada de Fourier de la HRIR (*Head-Related Impulse Response*). Para obtener el HRIR, el cuál no es más que un impulso medido al final del canal auditivo, se coloca un micrófono dentro de cada oído del sujeto y se le somete a una serie de impulsos auditivos desde varias direcciones. Dichos micrófonos grabarán el sonido filtrado por su cabeza, torso y oreja, y por comparación entre la señal emitida y la señal captada por los micrófonos se obtiene la respuesta HRTF personal del sujeto. Cuantos más impulsos desde un mayor número de direcciones y planos, mejor será la HRTF y más realista y cercana a la realidad será su respuesta.

Otros indicios

Por si todos los mecanismos utilizados por el cerebro mencionados hasta ahora no fueran suficientes, existen otra serie de mecanismos para la localización espacial de fuentes sonoras, intentando solventar las carencias de los métodos anteriores, y sirviendo de apoyo en caso de que éstos no fuesen suficiente.

Indicios dinámicos: Los indicios dinámicos no son más que un recurso utilizado de manera involuntaria a la hora de intentar localizar una fuente que no tenemos claro donde está. Cuando una fuente sonora emite fuera de nuestro ángulo visual, en ocasiones ésta se puede encontrar en nuestro cono de confusión, provocando que no tengamos claro de donde procede el sonido. Es en estos casos cuando, de manera involuntaria, tendemos a girar la cabeza e incluso el torso completo en busca de un apoyo visual, unas diferencias binaurales diferentes, e indirectamente, obtener otro ángulo de escucha con una HRTF distinta, la cuál pueda aportarnos mayor información que en la posición original, pudiendo salirnos así del cono de confusión y obteniendo incluso una mejora en la percepción de la elevación de la fuente.

Todo esto quedó reflejado por *P.T. Young* (1931), quien sometió a varios sujetos a estudios científicos donde sustituyó la pinna de los mismos por una especie de embudo neutro. Eliminando los pliegues y efectos acústicos producidos por ésta, los sujetos tendían a mover la cabeza en busca de la fuente sonora para compensar las carencias de localización provocadas por la falta de la pinna.

Inicialmente puede no parecer un fenómeno especialmente relevante para la localización, pero es una fuerte limitación a la hora de trabajar con audio binaural, ya que si el sujeto está escuchando un espacio sonoro con auriculares y una fuente no termina de localizarla con exactitud, instintivamente girará la cabeza y el torso como hace en la realidad para buscar dicha fuente, la cuál permanecerá invariable. Para solucionar este problema, se desarrolló una técnica denominada *head-tracking*, la cuál consiste en capturar los movimientos reales de la cabeza del sujeto con sensores de movimiento y, a través de una serie de algoritmos basados en binauralidad, calcular los fenómenos de indicios dinámicos entre otros para que el sujeto perciba las variaciones provocadas por el movimiento de su cabeza. De esta manera, se eliminarán varios conflictos de posición de las fuentes, como la internalización y las inversiones, como ya se ha visto anteriormente.

Esta tecnología de *head-tracking* es ampliamente utilizada en el entorno de los videojuegos, VR y AR.[5]

Efecto Doppler: El Efecto Doppler, denominado así por *C.A. Doppler* (1842), determina que una fuente sonora en movimiento genera un frente de ondas más próximos, y por tanto de mayor frecuencia a la original por delante de la dirección hacia la que avanza, quedando un frente de ondas de menor frecuencia justo por detrás, tal y como se observa en la figura 2.8.

Se trata del clásico ejemplo producido por una ambulancia, que a medida que se nos acerca y se nos aleja, la frecuencia de la misma cambia por el movimiento de ésta. Otro ejemplo se da en las carreras de automovilismo, donde el coche emite un sonido de frecuencia variable a medida que pasa por nuestro punto de escucha. La velocidad del mismo, provoca que nuestra percepción del sonido que emite varíe por el efecto comentado.

El Efecto Doppler es importante para la localización espacial de fuentes sonoras en movimiento, aportándonos información tanto de por donde se acerca o aleja la fuente, como incluso nos puede dar información sobre si la fuente se desplaza rápida o lentamente.

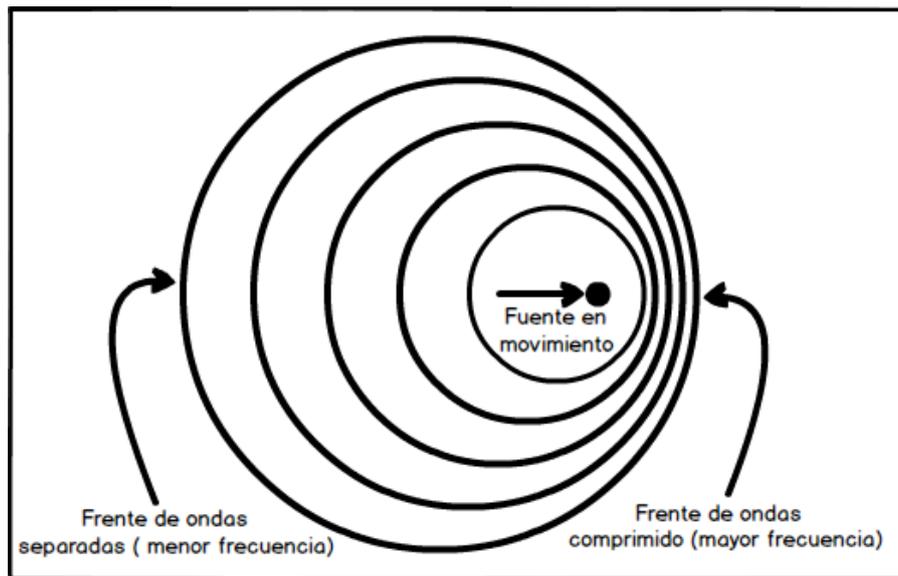


Figura 2.8: Efecto *Doppler*.

Efecto Cocktail Party: Otro de los fenómenos que nos ayudan a localizar fuentes sonoras es el conocido como Efecto Cocktail Party, definido y bautizado así por *C. Cherry* (1953). El sistema auditivo humano nos permite localizar y “enfocar” una fuente entre varias similares a ella dentro de un espacio sonoro. Dicho de otra manera, podemos encontrarnos en una reunión o un bar y, con todo el mundo hablando a la vez, somos capaces de centrarnos en una voz en concreto y escucharla, aunque todas compartan la misma banda de frecuencia incluso.

Es importante destacar que este efecto funciona mucho mejor con personas con una escucha binaural sana, por lo que podríamos decir que es un indicio binaural. Ya que con una audición deficiente en alguno de los dos oídos, el cerebro no es capaz de discriminar con la misma eficacia los sonidos que no nos interesan. Además, el apoyo visual de la fuente también ayuda mucho a la focalización del sonido deseado, volviéndose costosa si tenemos un apoyo visual nulo. Se trata de un efecto que tiene que ver más con el cerebro y cómo procesa la información, que con la audición propiamente dicha, ya que no se trata de un indicio únicamente auditivo, si no que necesita de un apoyo visual para que se dé correctamente.

Reverberación: La reverberación no es un indicativo al uso para la localización espacial pero nos ayuda a hacernos una idea de, por ejemplo, el tamaño de la sala o los materiales de la misma. Así mismo, una reverberación con una cola muy larga puede crear confusión a la hora de ubicar una fuente sonora, tal y como ocurre en iglesias y catedrales, donde en muchos casos, la inteligibilidad de la palabra o la ubicación de la fuente se vuelve confusa y difícil.

Así mismo, existen cámaras anecoicas, donde sin un apoyo visual, y ante la falta de reflexiones, las cuáles se dan de manera natural en la naturaleza, también es complicado en muchas ocasiones adivinar por ejemplo la distancia de la fuente hasta nosotros. Esto es debido a que el cerebro humano, de manera evolutiva se ha adaptado al mundo en el que vivimos, donde existen reflexiones de manera natural. Cuando eliminamos de manera forzada éstas, el cerebro pierde varios de sus indicios a la hora de situar espacialmente la fuente, y por tanto, nos es difícil situarla en el espacio, tanto su distancia como su posición.

Presbiacusia: Otro de los factores que afecta a la localización espacial de los sonidos es la edad. Con ella aparece la presbiacusia, es decir, la pérdida progresiva de las altas frecuencias en la audición por causa de la edad.

Si bien no existe un consenso total entre la comunidad científica respecto al espectro audible humano, ya que existen autores que defienden que el espectro audible de un ser humano sano es de 20 Hz a 20 kHz, mientras que otros autores sostienen que en realidad es desde los 16 Hz hasta los 18 kHz. En cualquier caso, a medida que nos hacemos viejos, sufrimos una pérdida en la audición, concretamente en las frecuencias altas. Lo habitual suele ser una pérdida de 2 kHz desde las altas frecuencias por cada 10 años a partir de los 20 años. Es por tanto, que se puede afirmar que la edad es por tanto otro motivo más para una correcta localización espacial, ya que si el sujeto no recibe información en altas frecuencias por culpa de la presbiacusia, es muy probable que tenga dificultades a la hora de localizar correctamente una fuente sonora frente a otra persona que tenga una audición sana y sin pérdidas.

2.2 Audio multicanal: Los inicios del audio espacial

Si nos adentramos en la historia del audio previa al nacimiento de la estereofonía, es fácil detenerse en 2 años y 2 inventores concretos, principales responsables del nacimiento del audio tal y como lo conocemos, y más concretamente, de los primeros sistemas de captación, grabación y reproducción.

En primer lugar llegó el nacimiento del **fonógrafo**, inventado por *T.A. Edison* (1877), o eso se dice. En cualquiera de los casos, fue el primer sistema “asequible” y de producción comercial de la historia. El fonógrafo es un sistema que permitía la grabación y la reproducción de material sonoro en unos cilindros de metal recubiertos con algún material semi blando que permitiera grabar en el los surcos de sonido, como ceras a base de parafina. Permitía grabar entre 2 y 4 minutos de audio en cada cilindro. Inicialmente, con los materiales utilizados en los primeros años, la grabación se iba degradando al cabo de unas decenas de reproducciones, por lo que a medida que la tecnología fue madurando, en torno a 1906, se fueron sustituyendo las ceras por otros materiales más permanentes, haciendo las grabaciones fonográficas casi indestructibles. [6]



Figura 2.9: Fonógrafo de la marca Edison.[6]



Figura 2.10: Cilindro de cera negra de la marca Edison de 1904.[7]

El siguiente paso lo dio *E. Berliner* (1887), quién inventó el **gramófono**. El gramófono es un sistema de grabación y reproducción analógico basado en los famosos discos de vinilo. Al principio nació como competidor directo del fonógrafo, ganando este en algunos apartados técnicos y perdiendo en otros, como la resistencia de los cilindros. En 1910, y tras mucha presión a base de campañas

de publicidad y grabaciones exclusivas por parte de las discográficas, el gramófono se antepuso al fonógrafo como sistema de distribución estándar para grabaciones.



Figura 2.11: Gramófono con bocina de amplificación.[8]



Figura 2.12: Disco de vinilo de 12" a 33 RPM.[9]

2.2.1 Estéreo

Théâtrophone

Justo unos años después de la invención del fonógrafo, y unos antes de la llegada del gramófono, *C. Ader* (1881) inventó el **Théâtrophone**, el primer sistema de captación y reproducción de audio de dos canales basado en el sistema telefónico. El théâtrophone fue instalado por primera vez para la captación de la Ópera de París y su reproducción en la Exhibición Eléctrica de París de 1881. Se colocaban un *array* de micrófonos en el escenario con la idea de capturar todo el espacio sonoro, el cuál se transmitía por cable telefónico hasta unos módulos personales equipados con dos auriculares (uno para cada oreja). El usuario podía estar en un espacio alejado de la ópera y aún así poder escucharla con una imagen estéreo bastante sofisticada para la época, siendo un invento que se asentó como método de difusión informativa. Existieron otras variantes del théâtrophone, como el **Telefon Hírmondó** en Budapest, Hungría en 1893, o el **Electrophone** en Londres, Reino Unido en 1895.

Finalmente, el théâtrophone fue perdiendo adeptos como sistema de difusión informativo y musical, dejando paso a sistemas más sofisticados como la radio o el gramófono. Cesó sus operaciones en 1932, dejando como legado ser el primer sistema de reproducción de dos canales, si bien todavía el concepto estereofonía estaba lejos de alcanzar su madurez.

Alan Blumlein y la llegada del estéreo

Tras unos inicios del sonido estéreo tan rudimentarios con el téâtrophone como protagonista, *Alan Blumlein* (1933), ingeniero de sonido de EMI, patenta el primer sistema de grabación estéreo, el cuál lleva su nombre. Dicha técnica consiste en la colocación de dos cápsulas bidireccionales coincidentes formando un ángulo de 90°, figura 2.15, siendo a día de hoy una de las técnicas de captación estéreo más utilizadas por su buena sensación espacial y su captación uniforme de la reverberación natural de la sala.

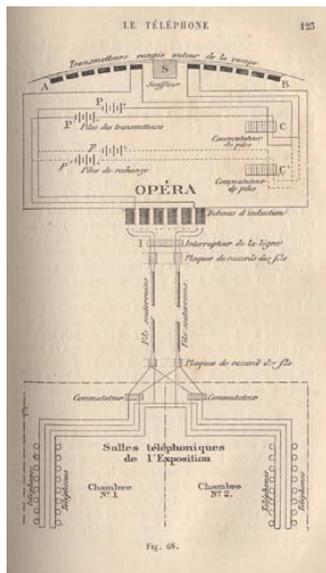


Figura 2.13: Diagrama del prototipo del Théâtrophone de 1881.[10]

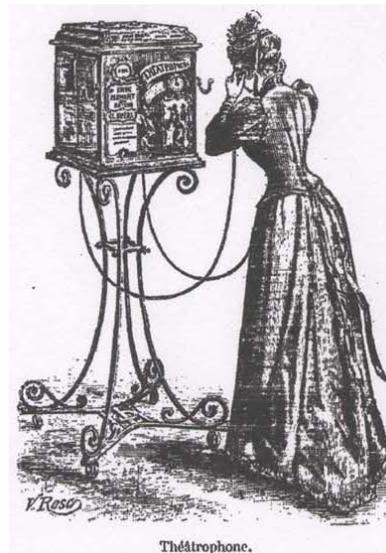


Figura 2.14: Ilustración de un Théâtrophone de *Le Magasin Pittoresque* (1892).[10]

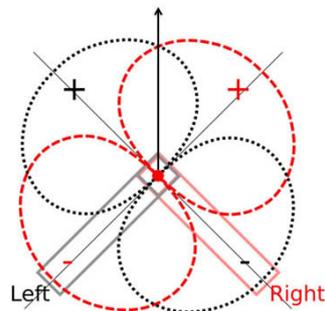


Figura 2.15: Principio de la técnica de captación estereofónica Blumlein.[11]

Paralelamente al trabajo de Blumlein, los Bell Labs, con *Harvey Fletcher* (1933) a la cabeza, investigaron su propia versión de audio estereofónico. Una de esas versiones se denominó “muro de sonido” y consistía en un conjunto de unos 80 micrófonos colocados frente a la orquesta en diferentes posiciones, y un conjunto de altavoces colocados en posiciones idénticas a los micrófonos en otra sala para la reproducción, de manera que reconstruyera el campo sonoro original.

Ese mismo año, los Bell Labs hicieron una demostración de grabación binaural en la Chicago World’s Fair mediante el uso de una *dummy head*² enviando ambas señales separadas por AM entre dos estaciones separadas.

Otra de las pruebas relevantes fue cuando el 27 de abril del 33 los Laboratorios Bell realizaron una demostración de sonido estereofónico basado en 3 canales con transmisión en vivo de la Orquesta

²Dummy head.- Sistema de grabación binaural que utiliza un prototipo de cabeza humana con un micrófono en el interior de cada oreja, recogiendo así todos los efectos acústicos de la escucha humana, como reflexiones, retardos, diferencias de fase, etc.

de Filadelfia, siendo la transmisión desde Filadelfia hasta el Constitution Hall de Washington D.C. a través de varias líneas de la red telefónica. Cinco años más tarde, se utilizaría por primera vez el mismo sistema para la grabación multicanal de audio para la película *Fantasia* (1940), siendo la transmisión desde la sala de conciertos de Filadelfia hasta los estudios de grabación de los Laboratorios Bell en Nueva Jersey. A esta tecnología, Disney la bautizó como *Fantasound*.^[12]

2.2.2 Cuadrafónico

A medida que el sistema estéreo se convertía en estándar y se iba adaptando a los diferentes medios de difusión y reproducción como la radio, el cine o la música, se cuestionaban las distintas limitaciones técnicas del mismo, como su baja cobertura espacial, con no más de 60° horizontales. Esta búsqueda de nuevos formatos fue la causa del nacimiento de los primeros sistemas de audio envolvente multicanal.

Para llevar a cabo dichas ideas, se probó a aumentar el número de altavoces de un sistema, el número de canales del mismo, extender la información direccional de un sistema estéreo desde los 60° mencionados hasta los 360° , cubriendo así todo el plano horizontal, o incluso cubrir todo el espacio sonoro incluyendo también el plano vertical.

Para ello, se evolucionó el sistema estéreo convencional, también conocido como 2.0 por su composición en el número de altavoces y canales, hasta el sistema **cuadrafónico**, siendo un 4.0 si seguimos con la misma equivalencia, y el cuál estuvo vigente entre 1969 y 1980.

Fue el primer sistema de sonido envolvente producido de forma masiva y comercial para el gran mercado, tanto a nivel profesional como doméstico, y precursor de sistemas de audio envolvente más sofisticados como *Dolby Surround*.

Un sistema cuadrafónico consistía en 4 altavoces colocados a 90° entre sí, formando un cuadrado perfecto, y donde el oyente debía colocarse justo en el centro del cuadrilátero (*sweet-spot*). Cada altavoz reproducía un canal propio y diferente del resto.

Ciertas limitaciones técnicas, como la mala respuesta espacial del sistema cuadrafónico, las dificultades que tenían salas de cine, de música, etc, para adaptar sus equipos a la cuadrafonía hicieron que ésta no ganase los adeptos que se pretendía en un principio y la relegaron a un plano muy secundario, con un apoyo muy discreto tanto por parte de los consumidores, como por parte de los fabricantes de equipos, y por productoras y distribuidoras de cine y música.

Esto hizo que su paso por el mercado fuera fugaz y su fama mediocre, ya que fue un sistema complicado de implementar y que además ofrecía una respuesta espacial bastante pobre, con una serie de limitaciones en sus respuesta que hicieron de él un sistema con una vida muy corta.

2.2.3 Dolby y los sistemas multicanal

Tras los varios intentos por implantar diversos formatos de audio envolvente multicanal, y con la empresa Dolby Laboratories® bien posicionada en el mercado gracias a las tecnologías de restauración y codificación de audio que venían desarrollando desde 1966, en 1982 se presenta al mercado la tecnología Dolby Surround.

La tecnología *surround*, ofrecía una cobertura de 360 grados en el plano horizontal del oyente, situando a éste en el centro, y con una resolución espacial y un *sweet-spot* a la altura de un sistema comercial, no como ocurría con sistemas anteriores como el cuadrafónico, cuyas características acabaron relegándolo al olvido por su falta de calidad en muchas de ellas. Una de las características de la tecnología *surround* es que el número de altavoces o canales no define la tecnología en sí. Esto

quiere decir que, para el mercado doméstico, Dolby ofrecía soluciones modestas basadas en sistemas 5.1 o 7.1, configuraciones más que suficientes a nivel de canales/número de altavoces para un salón o una pequeña sala de proyecciones doméstica. Sin embargo, y bajo la misma tecnología, existía la posibilidad de un aumento de canales o número de altavoces para grandes salas de cine profesionales. Si bien es cierto que la tecnología *surround* ofrecía varias configuraciones, todavía se trata de una tecnología discreta en su configuración de canales, es decir, el número de canales define el número de altavoces.

Para aclarar esto podemos observar la figura 2.16. Tal y como observamos, el número de canales de audio que salen del procesador, es amplificado independientemente por un canal de potencia dedicado y se distribuye por la sala a los altavoces asociados a dicho canal. Esto quiere decir que, a pesar de que 3 altavoces puedan ser alimentados por la misma señal de audio, como por ejemplo los altavoces traseros Lss, Lrs, Rrs o Rss, únicamente llevarán asociados un canal de audio, y por tanto los 3 reproducirán la misma señal. Esto al final es una limitación por ser un tipo de sistema de audio envolvente discreto, donde el número de canales de audio define el número de posiciones y por tanto limita el número de altavoces independientes. Como ya veremos, esta limitación viene solucionada por sistemas de audio envolvente más sofisticados como Dolby ATMOS® o Ambisonics.

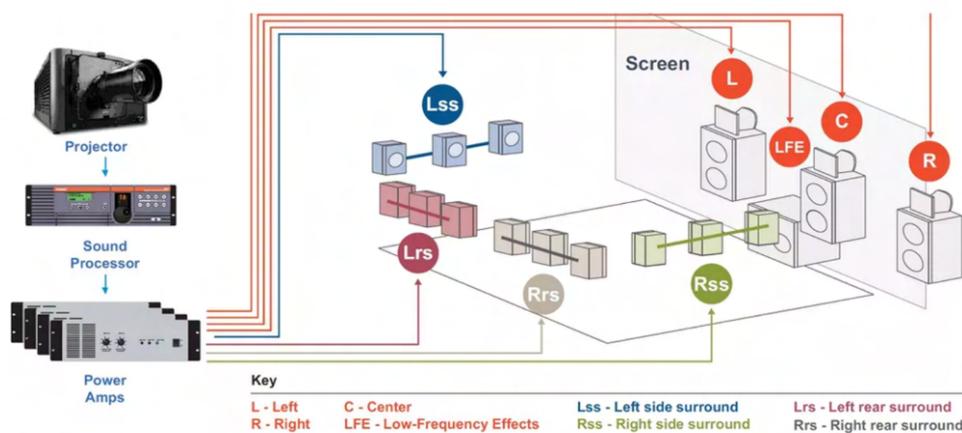


Figura 2.16: Ejemplo de sistema Dolby Surround 7.1 para salas de cine.[16]

Dolby ATMOS

En 2012 Dolby sacó al mercado su tecnología ATMOS. Por fin, y tras varios años en los que los formatos de audio envolventes consistían en sistemas discretos, en donde el número de canales nos indica la resolución espacial del sistema, es decir, a mayor número de canales, mayor número de altavoces y mayor resolución espacial. Con ATMOS se entraba en un nuevo paradigma de sistema de audio envolvente basado en objetos.

Los objetos, tal y como Dolby los bautizó, son fuentes puntuales de sonido que se mueven dentro de un campo sonoro. Para ello y durante la mezcla, el ingeniero no debe decidir por qué altavoz sonará según que fuente, si no que situará la fuente en la posición deseada dentro del campo sonoro y será el decodificador ATMOS quien decidirá por donde debe sonar dicha fuente. En otras palabras, será el decodificador quien, basándose en la *setup* de altavoces configurada, decidirá por que altavoces sonará la fuente para que la identifiquemos en la posición deseada.

ATMOS por tanto se compone de un sistema con una mezcla única para toda su distribución, tanto

para grandes salas de cine como para sistemas domésticos. Será el usuario final (bien sea doméstico o profesional) quien en su sistema, configurará el decodificador ATMOS con su sistema de altavoces, y éste decodificará cada objeto para que suene por la posición deseada. Es por ello que el número de canales de audio de la mezcla no determinará la resolución de la misma, ya que la mezcla de audio en realidad contiene un campo sonoro 3D completo, y éste se decodifica para cada plantilla de altavoces y/o auriculares. Dentro del formato ATMOS, además del audio, se encapsulan una serie de metadatos con la información relevante a cada objeto y a su posición dentro del campo sonoro.

Habiendo visto a *grosso modo* la tecnología ATMOS, se puede observar cómo Dolby se acercó mucho a la idea en la que se basa Ambisonics, donde lo que se codifica y distribuye es un campo sonoro 3D completo y universal, y es en la decodificación donde adaptaremos dicho campo sonoro a nuestro sistema de escucha deseado (2.0, 2.1, 5.1, 7.2.4, binaural, etc).

Si bien es cierto que ATMOS, al igual que Ambisonics, se puede decodificar para adaptarlo a sistemas domésticos de audio como el 2.1 o el 5.1, la verdadera potencia del sistema radica en los altavoces superiores, es decir, en la capa que corresponde a la altura. Para ello, el fabricante recomienda encarecidamente la colocación de altavoces en el techo para dar al oyente continuidad en los efectos de audio que pasen por encima. Esto está muy bien para salas de cine donde la instalación de sistemas en el techo se vuelve viable, pero se complica para sistemas domésticos. Además, hay que añadir la dificultad que supone el trabajar con sistemas de audio sin compresión de tantos canales en entornos domésticos. Por otro lado, y en caso de que nos guste el contenido en español, ya sea original o doblado, se vuelve muy complicado encontrar contenido que se distribuya con audio ATMOS. Esta limitación cada vez va siendo menos importante debido a todo el nuevo contenido y remasterizaciones que se lanzan al mercado con ATMOS entre sus características.

Es por tanto, que en países como España, la tecnología Dolby ATMOS ha quedado relegada más al ámbito comercial, para salas de cine y teatros, que para el mercado doméstico. Si bien es cierto que se han lanzado barras de sonido, auriculares y hasta smartphones con la certificación ATMOS de Dolby, éstos quedan muy lejos de la experiencia real que es capaz de ofrecer un sistema ATMOS completo, y por tanto se trata más de una estrategia de marketing que de un sistema ATMOS real.

Capítulo 3

Ambisonics: Teoría y Fundamentos

Ambisonics es una tecnología que nace en la década de los 70 en los laboratorios de la NRDC¹ británica bajo la batuta del matemático *Michael Gerzon*, ideólogo de Ambisonics. Dicha tecnología permite **grabar, procesar, distribuir y reproducir** audio espacial en 3D real en un formato totalmente escalable y compatible con el resto de sistemas de audio espacial, estereofónicos o monoaurales. Para llevar a cabo dicha tarea, Ambisonics se basa en la teoría matemática de los armónicos esféricos, figura 3.1, desarrollada por el propio Gerzon, la cuál no es si no un M/S² ampliado mediante dos señales más para obtener información tanto en altura como en la parte trasera del oyente, consiguiendo con un mínimo de 4 señales sonoras la representación acústica de una esfera de 360°. Es decir, Ambisonics no posee una señal para alimentar cada altavoz, como en el caso de los sistemas Dolby 5.1, 7.1, etc. Ambisonics codifica el campo sonoro completo de 360° en un mínimo de 4 canales, los cuáles contendrán toda la información sonora y espacial del campo. Una vez tengamos esas señales, podremos decodificar en el sistema de reproducción que deseemos, incluidos sistemas mono, estéreo o binaural.

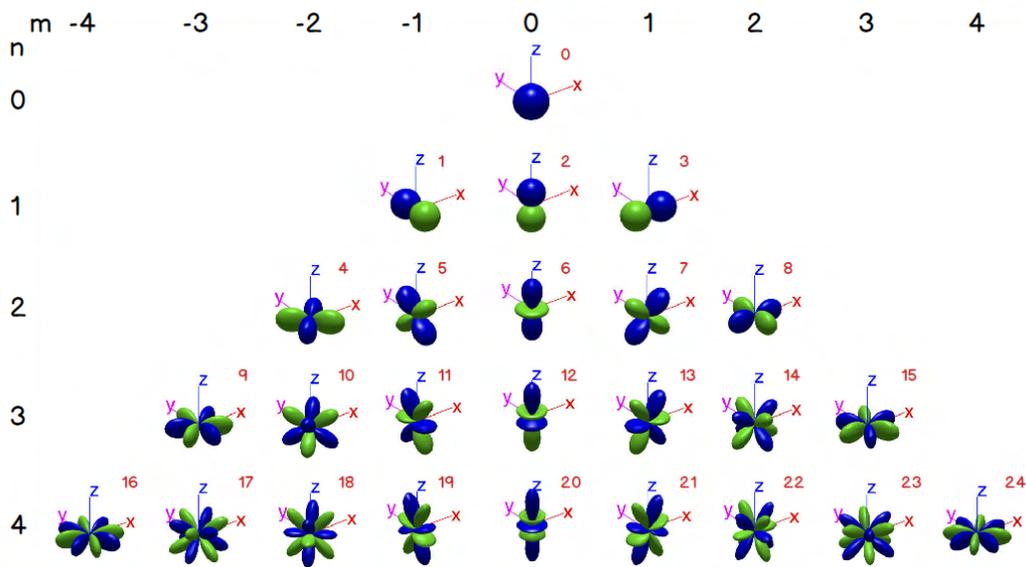


Figura 3.1: Armónicos esféricos hasta orden $n = 4$.^[17]

¹NRDC.- National Research Development Corporation.

²M/S.- Técnica Mid/Side de grabación en formato estéreo a partir de un micrófono cardioide encargado de la imagen central del estéreo y otro micrófono figura de 8 encargado de las señales laterales de la imagen sonora.

El hecho de que se diga que mínimo hacen falta 4 señales es por el hecho de que en Ambisonics podemos trabajar en diferentes órdenes, los cuáles vienen representados en la figura 3.1 por la letra n de la parte izquierda. En ella podemos observar que el orden 0 representa únicamente un micrófono omnidireccional sin ninguna información espacial. Si subimos a primer orden, harán falta 4 señales para la representación del campo sonoro completo, mientras que obviando la señal $m=0$ $n=1$ tendríamos un campo 2D completo, es decir, los 360° del campo horizontal. A medida que aumentamos el orden, la resolución espacial aumenta, pudiendo discriminar más direcciones dentro del campo sonoro pero necesitando un número mayor de señales para la representación de sus armónicos. Además, en color rojo tenemos el número ACN de cada armónico esférico, el cual se explica en la siguiente sección, y la expresión matemática que lo define se puede consultar en el Anexo **A**.

Por tanto, el número de canales necesarios para cada orden viene en la siguiente tabla, estando indicado tanto para una captación de 360° en 2D, es decir, sólo el plano horizontal, como el campo sonoro completo en las tres direcciones:

Orden	Número de canales	
	2D	3D
0	1	1
1	3	4
2	5	9
3	7	16
4	9	25
5	11	36
6	13	49
7	15	64

Antes de pasar a explicar cada uno de los procesos y subprocesos de Ambisonics, en la figura 3.2 podemos observar una cadena de trabajo genérica para trabajar con esta tecnología. En ella se puede observar cómo podemos partir de varios tipos de fuente: mono, estéreo, multicanal o incluso una grabación en formato Ambisonics desde un micrófono dedicado. Después podemos realizar toda nuestra mezcla en formato Ambisonics y tras esto decodificarlo al formato que más nos interese para su reproducción. En ello se puede ver la ventaja de que, una sola mezcla en Ambisonics permite ser decodificada en cualquier sistema de reproducción, por lo que no es necesario realizar una mezcla estéreo y otra en 5.1 por ejemplo para el caso del cine, ya que la distribución se realiza en un formato universal que contendrá toda la información del campo sonoro, siendo éste decodificado en base al sistema de reproducción elegido.

3.1 Grabación. Micrófonos Ambisonics.

La teoría en la que se basa Ambisonics, y si nos fijamos en la figura 3.1, podemos observar como para grabar en Ambisonics de orden 0, siendo n el orden, únicamente necesitaríamos un micrófono omnidireccional situado en el centro de la esfera que deseamos capturar. Como es lógico, un orden 0 capta el audio que le llega desde cualquier dirección pero no posee ninguna información espacial sobre ningún sonido. A esta señal correspondiente a un micrófono omnidireccional se le conoce como señal W.

Sin embargo, es cuando subimos a primer orden, también conocido como FOA (First Order Ambisonics), cuándo añadiendo 3 micrófonos con patrón figura de 8 situados sobre cada uno de los tres ejes de un sistema cartesiano (xyz), obtenemos información espacial sobre cada uno de los sonidos que nos

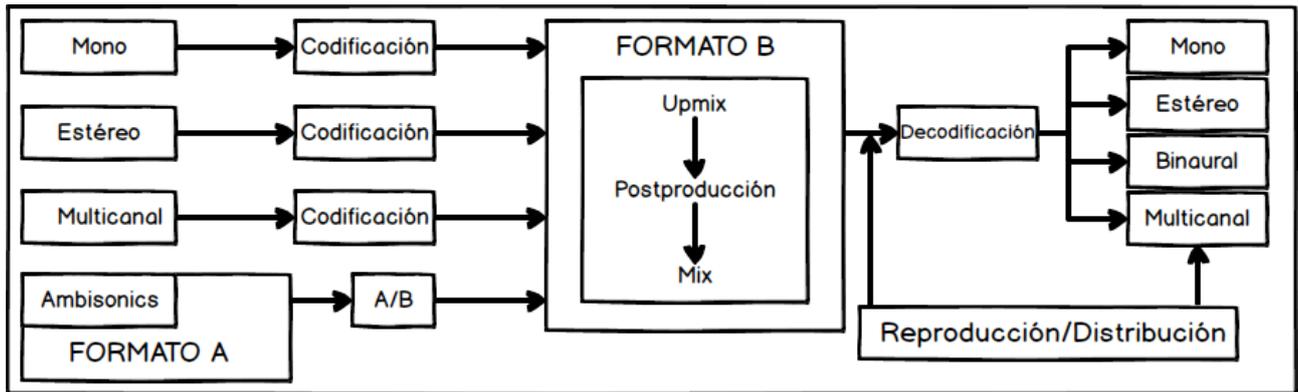


Figura 3.2: Flujo de trabajo dentro de Ambisonics.

llegan. Por tanto, con 4 señales, seremos capaces de obtener un campo sonoro de 360°. Es lógico pensar que si eliminásemos la señal $n=1 m=0$, obtendríamos un campo sonoro de 2D en el plano horizontal. A estas 3 señales correspondientes a cada uno de los tres micrófonos figura de 8 se les conoce como X, Y y Z.

Ahora bien, la teoría de los armónicos esféricos nos dice que todas las señales deberán ser capturadas desde el mismo punto central de la esfera de sonido para evitar problemas de fase entre las cápsulas. Esto se hace físicamente imposible, ya que no es posible alojar 4 cápsulas situadas en el mismo espacio central. Es por ello que a los fabricantes, con la marca Soundfield fundada por el propio Gerzon a la cabeza, ofrecieron una alternativa viable al problema de alojar las 4 cápsulas en la misma posición.

Si grabásemos directamente con un micrófono con una cápsula omnidireccional y 3 de figura de 8 situadas en la misma posición, éste nos daría 4 señales conocidas como **Format-B**. El formato B corresponde a las señales W, X, Y y Z antes mencionadas. Como esto no es viable físicamente, la solución pasó por alojar 4 cápsulas **cardioides** en las caras de un tetraedro, intentando que estén lo más cercanas al centro posible. A este formato se le conoce como **Format-A**.

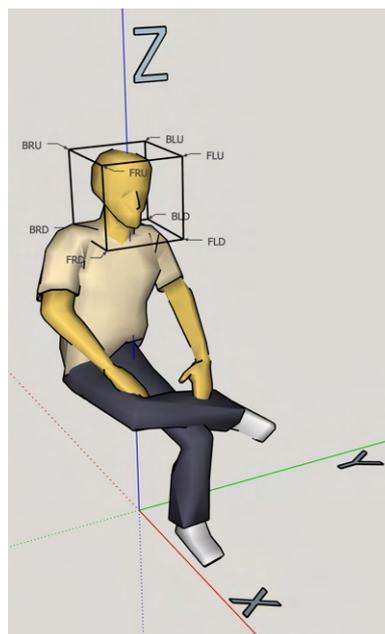


Figura 3.3: Sistema de coordenadas antropométrico según norma ISO 2631.

Como se observa, cada dirección viene indicada por la esquina donde convergen 3 caras del cubo, indicándonos mediante la nomenclatura de qué dirección se trata. La primera letra indica la posición delante-detrás (F.- Front/B.- Back). La segunda nos indica la posición izquierda-derecha (L.- Left/R.- Right). Por último nos indica la posición arriba-abajo (U.-Up/D.- Down). La situación de las cápsulas suele estar estandarizada por los fabricantes y la mayoría utilizan el mismo orden de colocación, siendo éste **FLU-FRD-BLD-BRU**. Existe una excepción, el micrófono prototipo de la casa *DPA*, conocido como *DPA-4*, cuyas cápsulas utilizaban las posiciones FLD-FRU-BLU-BRD, y del que se fabricaron muy pocas unidades.

3.1.1 Micrófonos HOA

Si bien se han explicado los micrófonos Ambisonics de primer orden, FOA, existen en el mercado una serie de micrófonos Ambisonics de órdenes superiores con diferentes variantes. Cada uno de ellos dispone de distintas peculiaridades, con alguna que otra limitación y/o ventaja que los hacen únicos. Para los micrófonos HOA, ya no se sigue una estandarización en la colocación de las cápsulas, ya que cada fabricante implementa sus tecnologías propietarias sobre la situación de las cápsulas y su conversión. Esta heterogeneidad se soluciona por el propio fabricante, el cuál se encarga de ofrecer (habitualmente) un sistema puente que convierta las señales captadas por su micrófono en formato B de orden superior para que el usuario pueda trabajar como de costumbre.

En primer lugar y con mayor veteranía en el mercado, existe un micrófono capaz de grabar en orden 4 mediante un *array* de 32 cápsulas conocido como *em32 Eigenmike®* de la casa *mhAcoustics*. Además, con fecha Noviembre de 2019, se acaba de presentar el próximo modelo de 64 cápsulas, el *em64*, capaz de grabar en formato Ambisonics de orden 6. La transmisión de audio se realiza en el primer caso a través de un formato propietario vía Ethernet hasta una caja de conversión y de ahí por Firewire al PC, y vía Dante© para el caso del *em64*. Ambos micrófonos se pueden observar en la figura 3.4, y comparten la peculiaridad de que la esfera en ambos casos tiene el mismo diámetro a pesar de haber duplicado el número de cápsulas, por tanto su comportamiento en altas frecuencias será el mismo debido a que los problemas de fase vendrán a ser los mismos debido a que la separación entre las cápsulas y el centro de la esfera es el mismo.



Figura 3.4: Micrófonos Ambisonics HOA. De izda. a dcha. *Eigenmike® em32*, *CoreSound OctoMic* y *Eigenmike® em64*.

Por otro lado existen micrófonos de orden inferior a los mencionados pero capaces de grabar en HOA con alguna peculiaridad, como el caso del *CoreSound OctoMic*, el cual podríamos decir que posee la capacidad de grabar en orden 1,5. El motivo de este número tan peculiar es que, para el plano horizontal es capaz de grabar en orden 2, mientras que para el eje Z (arriba y abajo) la resolución espacial es de primer orden, por lo que podríamos afirmar que su orden es 1,5, algo así como un orden mixto o híbrido.

Por último existen micrófonos de gamas más bajas enfocados a un público más *amateur*, los cuáles suelen venir en paquetes cerrados de hardware más software y todo preparado para que el usuario no tenga que hacer prácticamente nada, si bien la calidad de las cápsulas y las posibilidades que ofrecen estos micros queda lejos de los ya mencionados, son una muy buena solución para pequeñas empresas o aficionados que quieran introducirse en el mundo de la grabación en formato Ambisonics. Entre esta gama de micrófonos podemos encontrar la grabadora Ambisonics *Zoom H3-VR* de primer orden, el micrófono *Zylia Pro* de tercer orden o el micrófono *Voyage Audio Spatial Mic* de segundo orden.

3.2 Orden de canales y normalización.

Uno de los grandes temas de debate dentro del mundo de Ambisonics siempre ha sido la estandarización de un orden para los canales. A lo largo de su historia han surgido varias nomenclaturas para el orden de los canales, y para que el resto de procesos involucrados en nuestra cadena de trabajo con material Ambisonics pueda tener éxito, es importante definir bien estos órdenes, ya que en la mayoría de las ocasiones a la hora de procesar o trabajar con Ambisonics, deberemos ser nosotros los que indiquemos a los *plugins* y programas qué orden de canales estamos siguiendo, así como normalización y ponderación de los armónicos, lo cual se verá más adelante.[30]

Además, es posible omitir partes de la definición multipolar esférica para contenido que no posea una resolución uniforme en todas sus direcciones (conocido como orden mixto u orden híbrido), visto con el micrófono *CoreSound OctoMic*, las cuáles también podría ser necesario definir como tratamos esas componentes faltantes dentro de nuestro campo sonoro.

Para casos en los que nuestra transmisión sea “por cable”, entendiendo “por cable” como un enlace multicanal digital real o **virtual** dentro de un motor de procesamiento de audio, como por ejemplo un DAW, estas propiedades deberán coincidir explícitamente en ambos extremos del enlace, ya que habitualmente no existirá disposición alguna para el intercambio de metadatos que ayuden a identificar ninguno de estos parámetros. En el caso de intercambio de archivos, el problema es mucho menos grave ya que esta forma de intercambio ofrece mucha más flexibilidad al poder incluir metadatos en el propio archivo que ayuden a identificar lo que nosotros deseemos.

Sin embargo, y a pesar de todo, en la práctica sólo dos formatos son los más extendidos. El primero permite HOA de hasta orden 3 y se conoce como *Furse-Malham*, el cual se trata del orden tradicional de formato B, mientras que en el más moderno, el SN3D, se sigue el orden ACN, cuyo orden se puede consultar en el Anexo A. En ninguno de los dos casos existe ninguna ambigüedad en cuánto al orden, normalización, ponderación o polaridad, y se hace raro ver casos en los que falte algún componente. Por último, y de carácter más limitado, tenemos el formato N3D, el cual sigue también el orden de canal ACN.

3.2.1 Orden de los canales

El formato B tradicional, FuMa, y con orden WXYZ, únicamente contemplaba el orden 0 y 1 para Ambisonics. Como causa de la fuerte correspondencia entre los armónicos esféricos y las direcciones

a las que apuntaban los patrones polares de las cápsulas utilizadas en los micrófonos Ambisonics, se hizo necesario ordenar y nombrar dichos componentes de una maenra natural, de forma que se denominaron de la misma manera que los ejes en un sistema de coordenadas geométricas.[30]

El problema vino cuando se empezaron a desarrollar órdenes superiores, tanto a nivel matemático como con micrófonos HOA. El problema vino porque, tal y como se observa en la figura 3.1, los armónicos esféricos se disponen de forma simétrica alrededor de todos los armónicos situados sobre el eje z, correspondientes a $m = 0$, con los términos seno con $m < 0$ a la izquierda, y los términos definidos por cosenos y $m > 0$ a la derecha. Se puede consultar la definición matemática de cada uno de los armónicos en el Anexo A.

- **Furse-Malham.-** Para hacer del Furse-Malham un formato HOA se llevó a cabo una extensión del formato hasta hacerlo de tercer orden, siendo la nomenclatura de los canales para segundo orden RSTUV, y para el tercero KLMNOPQ. Como se observa, el orden comienza para cada orden desde su armónico $m = 0$, correspondiente con el armónico simétrico de rotación sobre el eje z, para luego ir saltando hacia los extremos a derecha e izquierda, dejando los componentes horizontales para el final, tal y como se observa en la tabla 3.1. Esto hace del sistema FuMa un sistema de ordenación poco intuitivo y un tanto confuso debido al orden de los armónicos, problema derivado principalmente de su planteamiento inicial y su posterior ampliación para adaptarse a los sistemas HOA de hasta orden 3.

			W ₀			
		Y ₂	Z ₃	X ₁		
	V ₈	T ₆	R ₄	S ₅	U ₇	
Q ₁₅	O ₁₃	M ₁₁	K ₉	L ₁₀	N ₁₂	P ₁₄

Tabla 3.1: Orden de canales para el estándar Furse-Malham.

- **ACN.-** Para sistemas HOA, y como forma de facilitar la nomenclatura y el orden de los canales se adoptó el sistema ACN (*Ambisonics Channel Number*), el cuál ha alcanzado un amplio consenso entre la comunidad, y el cuál viene determinado por el algoritmo $ACN = n^2 + n + m$, siendo las variables m y n las mismas que en la figura 3.1.

			0			
		1	2	3		
	4	5	6	7	8	
9	10	11	12	13	14	15

Tabla 3.2: Orden de canales para el estándar ACN.

El sistema ACN es el utilizado para SN3D y N3D, los cuáles se verán a continuación.

3.2.2 Normalización

Para una buena reconstrucción del campo sonoro, y que la imagen sonora se muestre de manera exitosa, con una buena relación entre la señal W y las componentes espaciales, XYZ, para un caso de primer orden, será conveniente realizar una normalización de las señales. Esta normalización será la encargada de que el balance de señal entre los armónicos esféricos esté ajustado para una buena reconstrucción del campo sonoro. Además, esta normalización será la encargada de que cada componente individual

no exceda una ganancia de 1, de manera que todas ellas queden normalizadas a dicho valor para su correcto tratamiento posterior.

Este paso, se realiza justo después de la conversión A/B, explicada en la siguiente sección, si bien se ha colocado aquí por tratarse de otro de los procesos en los que existen varios métodos y deberemos conocer en qué estándar o método estamos trabajando para así asignarlo en los *plugins* y programas que estemos utilizando para trabajar con señales Ambisonics.

- **maxN.-** Como ya se ha mencionado antes, el esquema de normalización *maxN* normaliza cada una de las componentes individuales para que ninguna de ellas exceda nunca una ganancia de 1.0, de manera que la fuente monofónica adquiera una apertura panorámica.

El esquema de normalización *maxN* es el utilizado en el formato clásico Furse-Malham, con la excepción de que aplica una reducción de ganancia de -3dB a la señal W, de manera que hace que éste sea compatible con el formato B clásico. Si esta atenuación de -3dB no es aplicada sobre la señal W, este formato no será estándar y quedará fuera de lo que comúnmente se conoce como esquema *maxN*.

- **SN3D.-** El significado de las siglas SN3D viene de semi normalización de Schmidt, técnica matemática utilizada habitualmente en geología y magnetismo. Dicha normalización viene definida por los coeficientes de ponderación derivados de la siguiente expresión:

$$N_{n,m}^{SN3D} = \sqrt{(2 - \delta_m) \frac{(n - |m|)!}{(n + |m|)!}}, \delta_m \begin{cases} 1 & \text{if } m = 0 \\ 0 & \text{if } m \neq 0 \end{cases}$$

La diferencia entre el SN3D y el N3D, el cual se verá a continuación, es que para SN3D ningún componente sobrepasará nunca el valor máximo de la componente W para fuentes de punto único.

SN3D con un orden de canales ACN es el estándar moderno actual más utilizado, por delante del Furse-Malham con normalización maxN, y dicha combinación de orden de canales ACN más normalización SN3D se conoce como formato **AmbiX**, con la diferencia de que AmbiX aplica una corrección a uno de los términos, tal y como aparece en la siguiente expresión:

$$N_{n,m}^{SN3D} = \sqrt{\frac{(2 - \delta_m) (n - |m|)!}{4\pi (n + |m|)!}}, \delta_m \begin{cases} 1 & \text{if } m = 0 \\ 0 & \text{if } m \neq 0 \end{cases}$$

- **N3D.-** Otro de los esquemas de normalización es el N3D, también conocido como normalización tridimensional completa, y es el enfoque más obvio para la normalización de señales Ambisonics. Su relación es la siguiente:

$$N_{n,m}^{N3D} = N_{n,m}^{SN3D} \sqrt{2n + 1}$$

Se trata de un tipo de normalización estándar y de sobra conocida en las ramas de física y matemáticas, y está soportada por la mayoría de los software y paquetes de Ambisonics, así como en estándar de vídeo como el MPEG-H. Sin embargo, en los últimos años ha visto como el esquema SN3D le ha comido terreno para ser el más común y utilizado.

Como los esquemas N3D y SN3D sólo se diferencian en los factores de escala, es importante tener cuidado al trabajar con ambos, ya que el cometer un error se hace difícilmente identificable de oído y puede ocasionar en errores y fallos a la hora de una buena reconstrucción del campo

sonoro, especialmente si tenemos material Ambisonics de varias fuentes o para sistemas con un número de altavoces reducido, debido a la baja resolución de éste.

- **N2D / SN2D.**- Por último, existen dos esquemas enfocados únicamente a las componentes horizontales. Esto es interesante porque ofrece ciertas ventajas prácticas en medios de procesado de punto fijo, especialmente en algunas situaciones bastante comunes donde las fuentes sonoras se concentran en el plano horizontal, para lo cuál la normalización se vuelve un tanto aleatoria y las suposiciones en las que se basa se vuelven inválidas para campo sonoro difuso o escenas sonoras con fuentes elevadas. Por tanto, y dado que Ambisonics es una tecnología isotrópica y que las escenas sonoras 2D realmente no lo son, dado que esto no puede darse en el mundo real, ya que inevitablemente tendremos reflexiones elevadas, etc, se desaconseja su uso.

3.3 Conversión A/B. Fuentes Ambisonics.

Una vez con nuestro material grabado y con las ideas claras acerca de los diferentes órdenes y estándares disponibles a la hora de trabajar con material Ambisonics, llega el momento de la conversión A/B. Para poder convertir de formato A a formato B, debemos realizar una conversión matricial con las siguientes fórmulas:

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

Si realizamos únicamente la conversión matricial expuesta, debido a la separación física de las cápsulas en el array, se nos producirán una serie de cancelaciones en alta frecuencia por la diferencia de fase con la que el sonido llega a cada una de ellas. Es por ello, que además de realizar las sumas y restas matriciales para obtener cada una de las 4 señales en formato B, deberemos aplicar un filtro de corrección a cada una de dichas señales antes de obtener el resultado final. Dicha corrección debe mantener la ecualización de cada señal pero corregir los problemas de fase en alta frecuencia. Dichos filtros vienen definidos por las siguientes expresiones[18]:

- Filtro W. Al aplicarlo sobre W', obtendríamos W:

$$F_W = \frac{1 + \frac{jwr}{c} - \frac{1}{3}\left(\frac{wr}{c}\right)^2}{1 + \frac{1}{3}\left(\frac{jwr}{c}\right)}$$

- r.**- Distancia de la cápsula al centro del tetraedro en metros.
- w.**- Frecuencia angular en rad/s.
- c.**- Velocidad del sonido en m/s.

- Filtro XYZ. Al aplicarlo sobre X', Y' y Z', obtendríamos X, Y y Z:

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

- r.- Distancia de la cápsula al centro del tetraedro en metros.
- w.- Frecuencia angular en rad/s.
- c.- Velocidad del sonido en m/s.

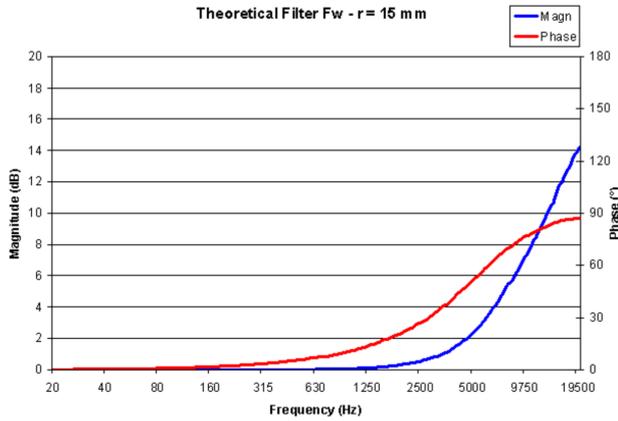


Figura 3.5: Respuesta teórica del filtro F_W en amplitud y fase para una separación $r=15mm$ entre cápsulas.[18]

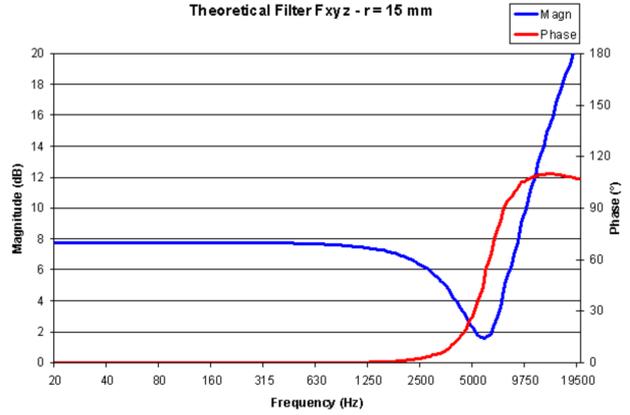


Figura 3.6: Respuesta teórica del filtro F_{XYZ} en amplitud y fase para una separación $r=15mm$ entre cápsulas.[18]

Tal y como se observa en la figura 3.5, tenemos un filtro nulo y lineal en baja frecuencia, zona del espectro donde no tenemos problemas por la estructura física de la colocación de las cápsulas. Sin embargo, cuando aumentamos en frecuencia, observamos como el filtrado consiste básicamente en un filtro tipo *shelving* positivo a partir de los 2500Hz aproximadamente, y con una ganancia de hasta 8-10dB en la zona alta del espectro, todo con la finalidad de corregir los problemas de cancelaciones que ocurren en esta zona del espectro por la separación de las cápsulas. Sin embargo, y tal y como se observa en la figura 3.6, para las cápsulas encargadas de aportar la información espacial, el filtro aplicado es diferente, con un aumento lineal desde la parte baja del espectro hasta los 2500Hz aproximadamente, donde se aplica un filtro tipo *peak* con una atenuación respecto al resto del filtro y por último un *boost* desde los 6000Hz en adelante. Cabe destacar que esto son únicamente filtros teóricos, por lo que al aplicarse sobre las señales procedentes de nuestra matriz de suma, se realizará una corrección bastante buena a la respuesta del micrófono que únicamente contemplará la variable r , es decir, la distancia de las cápsulas al centro del tetraedro.

Si bien la aplicación de este tipo de filtros es el método más común de corrección para micrófonos Ambisonics de primer orden, actualmente se están probando otro tipo de métodos, tal y como sugiere **Angelo Farina** en su artículo “*A-format to B-format conversion using X-volver*”, donde sugiere medir la respuesta al impulso de cada una de las cápsulas y, mediante un filtrado inverso, aplicar un filtrado de convolución a cada cápsula, de manera que la corrección deja de tener en cuenta únicamente la distancia de las cápsulas al centro y pasa a ser un filtro experimental adecuado a cada micrófono. Obviamente es un método mucho más personalizado pero más costoso, especialmente para fabricantes en masa. Existen marcas como *CoreSound* que entrega cada uno de sus micrófonos medidos y con el filtro de corrección experimental, siendo este un método mucho más fiable y con mejores resultados que el aplicar un filtro teórico, si bien éste es más sencillo y universal.

Una vez con las señales compuestas y filtradas, tenemos las 4 señales WXYZ correspondientes al formato B y listas para trabajar con ellas. Dicha conversión se puede realizar tanto a nivel hardware mediante módulos externos como el que incorpora el micrófono Soundfield ST450, o mediante software



Figura 3.7: Micrófonos Ambisonics de 1 orden. De izda. a dcha. *CoreSound Tetramic*, *Rode NT-SF1*, *Sennheiser Ambeo* y *Soundfield SPS200*.

como el SPS200 o el Rode NT-SF1 de la figura 3.7. El hecho de que la conversión se haga mediante un módulo externo o vía software tiene sus ventajas y desventajas. Entre las ventajas tenemos que a través del módulo ya podemos monitorizar vía auriculares el resultado en formato B, prescindiendo del uso de un PC o grabadora con conversor Ambisonics, y pudiendo grabar directamente en una grabadora de 4 pistas en formato B, muy recomendable para grabaciones en exterior. Sin embargo, el uso de software permite actualizaciones y mejoras en la conversión a medida que el fabricante las incorpore, pero en exteriores tendremos que grabar en formato A si no tenemos PC o como ya se ha mencionado, una grabadora con conversor Ambisonics incorporado, posibilitando así una correcta monitorización de la escena.

Además, y tal y como se ha visto, existen diferentes tipos de formato B dependiendo del orden de las señales o la normalización, si bien son dos los estándares comúnmente aceptados. Ambos formatos son el FuMa y el AmbiX. Esto se debe a que originalmente el formato más utilizado fue el Furse-Malham. Posteriormente, por mejoras en el flujo de trabajo y en el estándar entre fabricantes, se declaró el formato ACN (Ambisonics Channel Number), también conocido como AmbiX cuando lleva una normalización SN3D, como el estándar, siendo el orden de los canales el siguiente para cada caso:

- Furse-Malham (**FuMa**) = WXYZ

- ACN (**AmbiX**) = WYZX

Por tanto, es importante destacar la conversión de formato A a B como uno de los puntos críticos y de las limitaciones de Ambisonics. Cuanto menor sea la distancia r entre las cápsulas y el centro del tetraedro, mejor será la respuesta del micrófono en alta frecuencia. Debido a este conflicto en alta frecuencia este punto de procesado es un paso crítico en la respuesta del micrófono. Es por tanto fácil darse cuenta de porqué Ambisonics tardó tantos años en popularizarse comercialmente. La necesidad de un procesado de señal obligatorio e intrínseco al propio sistema, prorrogó su popularidad hasta que las capacidades de proceso de los ordenadores personales y los DSP³ fueron capaces de hacerse cargo de la conversión y normalización, ya que antiguamente ésta se realizaba de manera analógica

³DSP.- Digital Signal Processor.

mediante filtros bastante complicados a nivel de diseño, poco eficientes, y con resultados mediocres en la conversión.



Figura 3.8: VST del Rode NT-SF1 para conversión de formato A-B.



Figura 3.9: Box externo del Soundfield ST450 MKII para conversión y monitorización de formato A-B.

3.4 Codificación. Fuentes mono y estéreo.

Si bien podemos partir de una grabación Ambisonics nativa a partir de un micrófono dedicado a tal fin, otra de las opciones que tenemos es la de codificar cualquier fuente de audio en Ambisonics para trabajar en dicho formato durante la mezcla, al igual que haríamos con una fuente mono al colocarla en uno de los dos lados del estéreo. Para tal fin, podemos partir de fuentes mono, estéreo o incluso multicanal, y necesitaremos un *encoder* o *panner* Ambisonics. Lo habitual es trabajar con un *plugin* en formato VST dentro de nuestro DAW, el cuál se encargará de posicionarnos la fuente que elijamos dentro de nuestro campo sonoro. Para ello, el funcionamiento es similar a un *panner* tradicional pero con la posibilidad de colocar la fuente sobre la superficie de una esfera 360 grados.



Figura 3.10: *StereoEncoder* del Institute of Electronic Music and Acoustics.



Figura 3.11: *AmbiENC* de la suite SPARTA, desarrollado por el laboratorio acústico de la AALTO University.

En la figura 3.10 se puede observar como se ha codificado una fuente estéreo abriendo su imagen 90° sobre la frontal, es decir, situando su canal R a 45° y el L a -45°. Además, como se puede observar

en la parte superior, se ha codificado en orden 3, es decir, serán necesarios buses de 16 canales a partir de este punto para poder trabajar en Ambisonics. Por otro lado, en la figura 3.11, tenemos codificada una fuente 5.x según normativa. Esto quiere decir, que cada canal de altavoz del 5.1 se codifica en la posición que le correspondería dentro del espacio físico de la habitación. De esta manera, a partir de aquí podremos mezclar en formato Ambisonics y decodificarlo posteriormente como deseemos. Para este caso también se ha codificado en orden 3.

Como es lógico pensar, y como si de cualquier *panner* se tratara, existe la posibilidad de automatizar y programar cada uno de los controles del *encoder*, por lo que se podrá automatizar el movimiento de cualquier fuente dentro del campo sonoro. Por ejemplo, podremos automatizar dentro de nuestro DAW el sonido de un avión pasando por encima de nuestras cabezas, movimiento que no podemos llevar a cabo en una mezcla estéreo o 5.1 debido a que no tenemos la posibilidad de colocar sonidos en altura.

Por tanto, a partir de este punto, y una vez codificadas nuestras fuentes, ya podremos trabajar en formato B Ambisonics del orden elegido, pudiendo mezclar fuentes Ambisonics nativas con cualquier fuente codificada dentro de la misma mezcla. Un buen ejemplo podría ser un cuarteto de cuerda grabado con un micrófono en primer plano para cada instrumento y un micrófono Ambisonics en el centro para captar la respuesta natural de la sala. Después únicamente tendremos que codificar cada micrófono de instrumento en su posición real dentro de la sala y mezclarlo con el ambiente Ambisonics, consiguiendo así la naturalidad de la sala con cada instrumento en la posición que toca, y la respuesta precisa y brillante de un micrófono en primer plano.

3.5 Formato B. Upmix, postproducción y mezcla.

Con la codificación y/o conversión A/B, y estando ya con nuestro material de audio en formato B Ambisonics, se puede realizar ahora las tareas y procesos típicos de cualquier proyecto sonoro, siempre y cuando estos procesos se traten de procesos **lineales**, tales como aplicar ecualización o efectos temporales como reverb, delay, etc. La única e importante diferencia reside en que los procesadores VST/AU/AAX⁴ deben ser multicanal, es decir, ser capaces de trabajar para buses de más de dos pistas mono o una estéreo. Este detalle es una limitación la cual nos obligará a contar con menos procesadores de lo habitual, pero suficientes para una mezcla adecuada.

Antes de llevar a cabo ningún tipo de mezcla, y si buscamos aumentar el orden en el que estamos trabajando para alguno de los pasajes como por ejemplo las fuentes grabadas en formato A de primer orden, es posible aplicar un procesador conocido como *Upmixer*, el cuál nos aumenta de orden Ambisonics. De esta manera, y si por ejemplo queremos realizar toda nuestra mezcla en tercer orden, es posible codificar las fuentes mono, estéreo o multicanal directamente en tercer orden, y si queremos combinarlas con fuentes grabadas directamente en formato A o B, éste se puede aumentar de primer a tercer orden con esta herramienta, aumentando así su resolución espacial y adaptándola al orden del resto de fuentes. Es importante comentar que éste aumento de orden se realiza de manera matemática mediante interpolación, por lo que no deja de ser un aumento sintético, que si bien el resultado es muy convincente, no debemos olvidar que esa información extra no está captada por el micrófono durante la grabación.

Una vez situados en el orden que queremos, y con todas las fuentes convertidas, codificadas y adaptadas a nuestra mezcla, el flujo de trabajo que le sigue será similar a cualquier mezcla conocida. Tal y como se ha dicho, el único problema o limitación es el uso de procesadores multicanal para cada tarea. La automatización de las posiciones de las fuentes, así como cualquier parámetro que

⁴Diferentes formatos propietarios de plugins para operar dentro de un DAW.

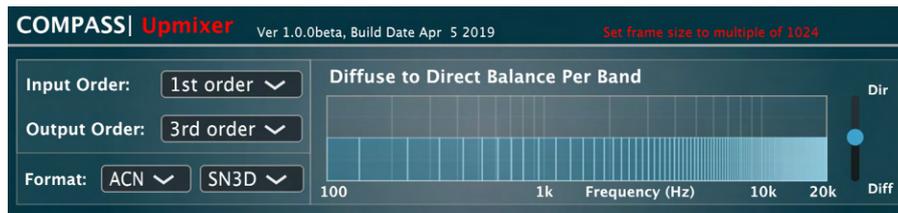


Figura 3.12: *Upmixer* de la suite SPARTA con una configuración de primer a tercer orden.

queramos automatizar se realiza igual que en cualquier mezcla, por lo que únicamente necesitaremos procesadores multicanal y un DAW que soporte buses de mezcla multicanal, como por ejemplo Reaper.

Además, se dispone de alguna herramienta de ayuda dedicada a poder mezclar en Ambisonics de manera cómoda. Por ejemplo existen visualizadores de energía para poder hacernos una idea de en qué zona del campo sonoro está sonando una fuente, su intensidad relativa respecto al resto de fuentes, etc. Así como un analizador de espectro en el dominio espacial en vez de frecuencial.

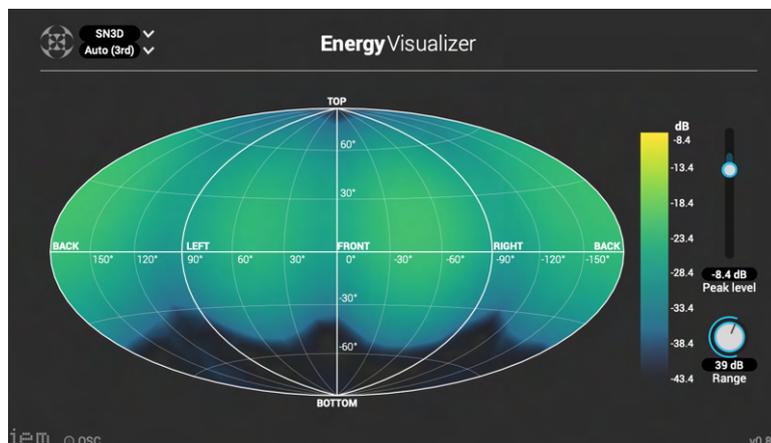


Figura 3.13: *EnergyVisualizer* de la suite IEM para tercer orden.

3.5.1 Aplicación de efectos no lineales en HOA mediante SPS

Habitualmente muchos de los efectos que queremos aplicar a nuestro contenido Ambisonics HOA los tendremos disponibles en las *suites* que se mencionan a lo largo del presente proyecto. Sin embargo, la mayoría de estos efectos son lineales, por lo que debido a su naturaleza lineal no presentan problema al ser aplicados a señales Ambisonics. Existen sin embargo otro tipo de efectos **no lineales** que un ingeniero o técnico de sonido querría poder aplicar, como por ejemplo efectos de dinámica tales como compresión, limitación, puerta de ruido, expansión, u otros efectos no lineales como la reducción de ruido. Para ello, algunos de estos efectos están disponibles en las *suites* mencionadas, pero otros no. Para ello, a continuación se propone un procedimiento para poder aplicar cualquier efecto de procesamiento de audio, lineal o no, a nuestro material Ambisonics de hasta tercer orden. El siguiente método está basado en el artículo de **Angelo Farina** llamado “*Performing not linear processing (Denoising, Compression, Limiter, etc.) on High Order Ambisonics signals using Adobe Audition CC and the SPS approach*”.[37]

Para llevar el siguiente procedimiento, utilizaremos como DAW *Reaper* y como paquete para trabajar en Ambisonics la *suite* SPARTA. A través de estas dos herramientas, convertiremos nuestro

material Ambisonics a formato SPS⁵. El formato SPS consiste en esencia en un formato formado por una serie de micrófonos virtuales cuyos patrones de directividad se colocan de forma que entre todos cubran uniformemente la superficie de una esfera. Como caso más básico tendríamos el de 8 micrófonos cardioides coincidentes apuntando cada uno a cada vértice de un cubo (disposición denominada Match1). Otro formato SPS sería el SPS-32, formado por 32 cápsulas cardioides de 4º orden colocadas en las mismas direcciones que las cápsulas del micrófono HOA *Eigenmike em-32*, tal y como se observa en la figura 3.14.

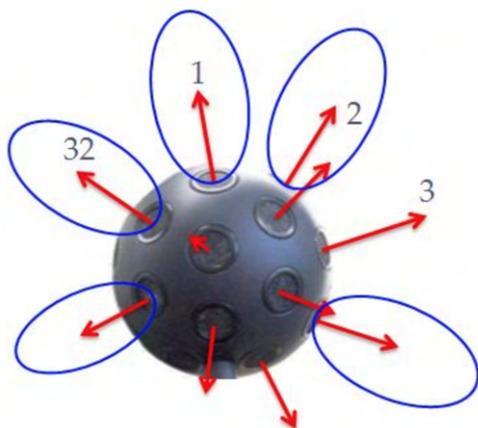


Figura 3.14: *Eigenmike em32* y la ordenación de sus cápsulas similar al SPS-32.[38]

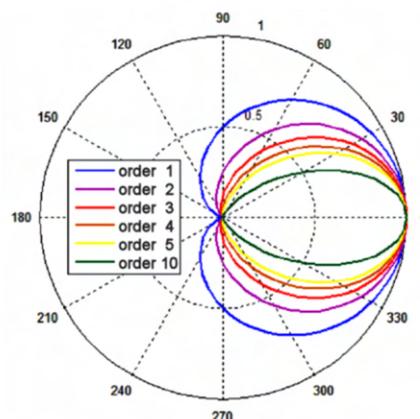


Figura 3.15: Patrón polar cardioide para diferentes órdenes.[38]

Para esos patrones cardioides utilizados en los formatos SPS mencionados, dichas señales SPS se conocen como “*P-format*” o formato PCM. Para ello, están formadas por patrones cardioides, como los que se observan en la figura 3.15, lo que hace que la señal capturada esté siempre en fase respecto a un micrófono omnidireccional (debido a que estos patrones no poseen lóbulos de polaridad negativa, como sí lo tienen patrones super o hiper cardioides, o figura de 8).

Como evolución del formato SPS basado en patrones polares cardioides, se viene utilizando una disposición mucho más uniforme de las direcciones de cada micrófono sobre la esfera virtual, proporcionando así una mejor cobertura gracias al uso de patrones polares virtuales multilobulares, así como con lóbulos de polaridad negativa obtenidos mediante el cálculo de las funciones SAD⁶, explicado en la sección 3.6.2. Se utiliza para calcular el número y las direcciones óptimas de los micrófonos virtuales sobre la esfera el método de “diseño en T”, por tanto, a este tipo de SPS se le conoce como “*T-format*”, dejando claro así que las direcciones vienen definidas por una distribución de las mismas en diseño en T.

El uso del formato T ofrece la posibilidad de realizar una conversión de señales Ambisonics HOA a SPS y retornar a HOA de manera totalmente reconstructiva y sin pérdidas siempre y cuando se utilicen el número suficiente de canales SPS. Por tanto, y gracias a este sistema, podremos realizar una conversión “*B-format*” > “*T-format*” > “*B-format*” siempre y cuando respetemos el número de canales en formato T para una correcta reconstrucción al volver a formato B. A continuación se expone una tabla con la relación entre el número de canales:

⁵SPS.- Spatial PCM Sampling.[38]

⁶SAD.- Decodificación Ambisónica de muestreo.

Orden Ambisonics	Canales “ <i>B-format</i> ”	Canales “ <i>T-format</i> ”
1	4	4
2	9	12
3	16	24
4	25	36
5	36	60

Tabla 3.3: Equivalencia de canales para formato B y formato T.

Para llevar a cabo dicha conversión entre formatos se van a utilizar los plugins de la *suite* SPARTA. En concreto se utilizarán el SPARTA AmbiDEC para la conversión “*B-format*” > “*T-format*” y el SPARTA AmbiENC para la conversión “*T-format*” > “*B-format*”.

Esto hace posible que, mientras en formato T los efectos no lineales como efectos de dinámica o eliminación de ruido no tienen consecuencias negativas sobre la información espacial del campo sonoro, preservando así la posición de cada fuente sonora dentro del campo, en formato B la aplicación de estos efectos puede desembocar en un completo desorden de la información espacial. Por tanto, la conversión de formatos, nos permite realizar un cambio de dominio para diferentes aplicaciones con el mismo material de audio espacial.

Para llevar a cabo el proceso comentado sin incidencias, a continuación se explica el flujo de trabajo dentro de *Reaper*:

1. Abrir la grabación en formato B.
2. Copiarla en un nuevo archivo WAV que contenga el número de canales requerido para su conversión a formato T. Consultar la tabla 3.3.
3. Aplicar SPARTA AmbiDEC para convertir a formato T.
4. Aplicar el efecto no lineal deseado sobre las señales en formato T.
5. Aplicar SPARTA AmbiENC para la conversión a formato B.
6. Deseleccionar los últimos canales vacíos y copiar los primeros en una nueva pista en formato B para que pueda ser guardada.

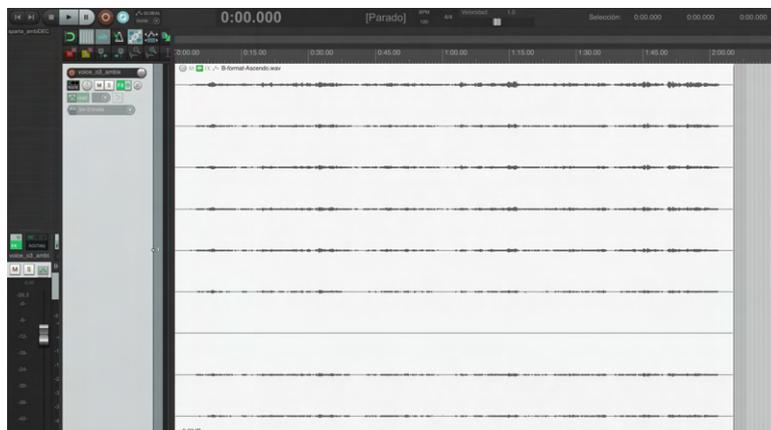


Figura 3.16: *B-Format* de 2º orden procedente de *OctoMic* de *CoreSound*.

En primer lugar, figura 3.16, debemos importar nuestra pista a *Reaper*. En nuestro caso se trata de una muestra descargada de la página oficial de *CoreSound* y realizada con un micrófono de segundo orden *OctoMic*. Como ya se ha mencionado, este micrófono no graba en segundo orden real, ya que únicamente cuenta con 8 cápsulas. Por tanto, y a pesar de que el archivo contenedor posee 9 canales, como si se tratase de una pista Ambisonics de segundo orden, observamos como el canal 7 se haya en blanco. Esto debemos tenerlo en cuenta a la hora de decodificar nuestro pasaje Ambisonics en la parte final (si necesitásemos llevar a cabo la decodificación).

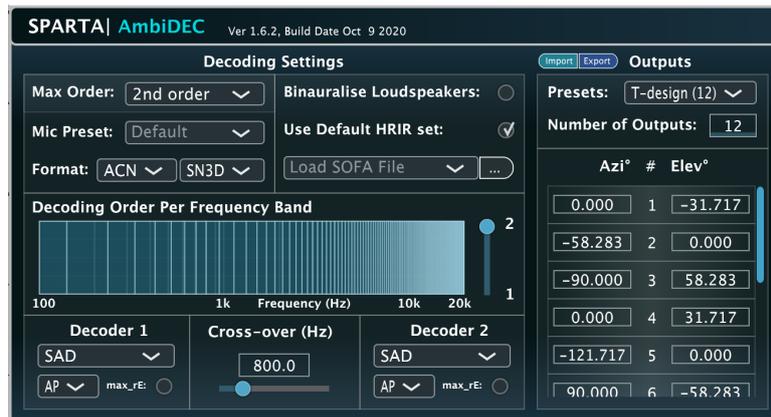


Figura 3.17: *AmbiDEC* configurado para convertir de 2º orden a *T-design* de 12 canales.

Una vez importado nuestro archivo y evaluadas sus características y puntualidades, es el momento de realizar la conversión “*B-format*” > “*T-format*”. Para ello, situaremos en el *rack* virtual de efectos el procesador *SPARTA AmbiDEC* con las especificaciones que se observan en la figura 3.17. Es importante prestar atención a que los decodificadores empleados sean de tipo SAD, ya que nos encontramos con un diseño en T, basados en este tipo de tecnología sobre patrones polares multilobulares, con lóbulos de polaridad negativa, obtenidos mediante el cálculo de las funciones de “decodificación ambisonica de muestreo” (SAD). El *plugin* nos incluye en los *presets* de salida plantillas para diseños en T, como el que se muestra, con la colocación de cada cápsula virtual según norma.

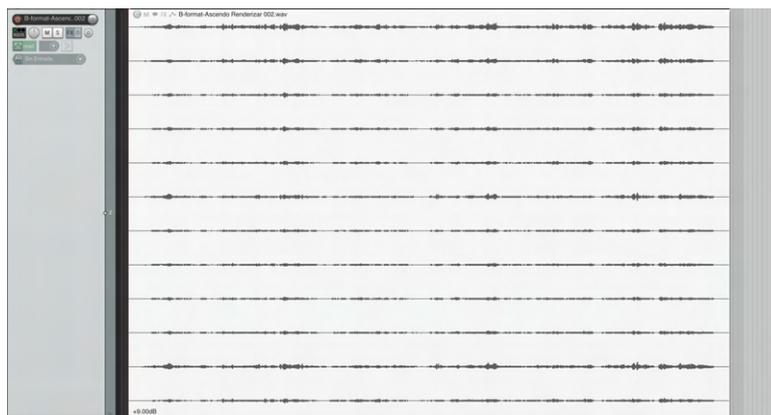


Figura 3.18: SPS en formato *t-design* de 12 canales.

Una vez configurado el *AmbiDEC*, realizamos un *bounce* de nuestro archivo para procesarlo y añadimos el archivo resultante a *Reaper*. Observamos, tal y como se ve en la figura 3.18 y como se adelantó en la tabla 3.3, cómo nuestro archivo en formato T posee 12 canales. Por tanto, en este punto nos encontramos en formato T y listos para realizar el procesado no lineal en nuestro archivo de audio.

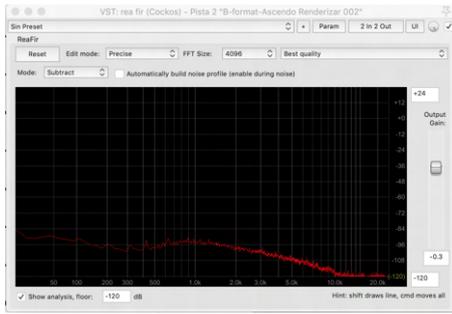


Figura 3.19: Perfil de ruido obtenido al final de uno de los canales de la grabación.



Figura 3.20: *AmbiENC* configurado para convertir de T-12 a Ambisonics 2º orden.

Para el ejemplo se decidió aplicar un *denoising* o reducción de ruido al material proveniente del *OctoMic*. En la figura 3.19 se observa el perfil de ruido obtenido en el silencio al final de uno de los canales de audio. Se trata de un perfil bastante bajo y apenas apreciable, pero sirve como ejemplo para documentar el presente proceso. Una vez obtenido el perfil en uno de los 12 canales, se aplica el efecto sobre todos ellos, asegurándonos de una correcta aplicación. Si el procedimiento ha ido bien, se debería haber eliminado parte del ruido inherente a la grabación a costa de sacrificar parte de la riqueza frecuencial del pasaje (consecuencia del procesamiento de reducción de ruido). Por este motivo deberemos ser cuidadosos con la aplicación de éste efecto.

Con el procesado aplicado, el siguiente paso es devolver nuestro material de audio espacial a formato Ambisonics de segundo orden. Para ello, se aplica en la *rack* de efectos el *plugin AmbiENC* con la configuración que se muestra en la figura 3.20. Se elige el orden destino y el preset que queremos aplicar, en este caso *T-design 12*.

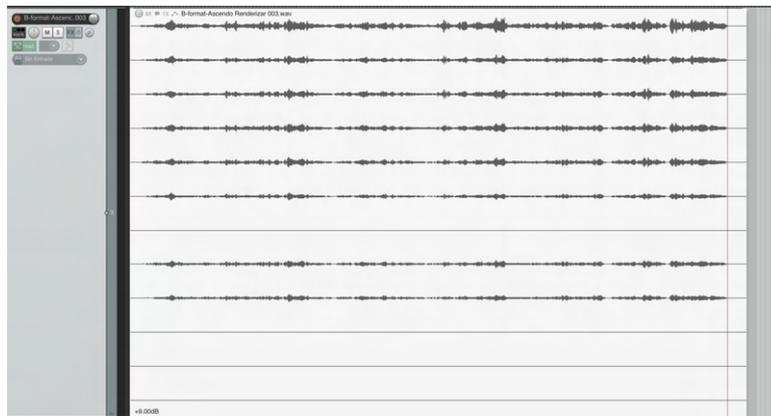


Figura 3.21: Resultado final procesado con 9 pistas *B-format*.

Por último, deberemos realizar un *bounce* final para procesar nuestro material y obtener un nuevo archivo Ambisonics en formato AmbiX de segundo orden. Si todo ha ido de forma satisfactoria, observaremos, como en la figura 3.21, un archivo con 12 canales, de los cuáles el canal 7 y los canales 10, 11 y 12 deberán estar en silencio. El canal 7 en silencio es un indicativo de que la conversión entre formatos ha ido bien, y por tanto tenemos al final un archivo similar al proporcionado por el *OctoMic*. Los tres canales en silencio del final se crean al utilizar buses de 12 canales cuando trabajamos en formato T, y no nos interesan al reconvertir a formato B, por lo que deberán ser eliminados.

Mediante el procedimiento explicado en la presente sección, podemos aplicar efectos tales como

compresión, limitación, eliminación de reverberación, eliminación de ruidos y restauración de audio, etc. Para la aplicación de estos efectos podemos utilizar los *plugins* incluidos en nuestro DAW, como en nuestro ejemplo, o de terceros.

Hay que prestar atención al procesado en pistas de formato B, ya que una aplicación de efectos no lineales en este formato puede alterar y destruir el delicado equilibrio ganancia/fase entre las señales Ambisonics de armónicos esféricos, destruyendo por completo la información espacial de éstos. En cambio, el formato T es mucho más robusto en este aspecto, ya que cualquier alteración de estos equilibrios únicamente lleva a una atenuación o amplificación de los sonidos procedentes de alguna dirección determinada, y no a la alteración de la información espacial del campo sonoro.

3.6 Decodificación y *playback*.

El último paso de nuestro flujo de trabajo se trata de la forma en la que queremos reproducir/ monitorizar y distribuir nuestro material Ambisonics. Esto es necesario porque por sí sola, una pista Ambisonics no es posible reproducirla sin decodificar⁷. Esto es, indicarle cómo queremos que descodifique el material.

Mientras que un archivo estéreo se puede reproducir tanto en auriculares, como altavoces éste no está pensado para ser reproducido por ejemplo en un *home cinema*, por lo que si lo hacemos nos encontraremos con que los altavoces traseros o central del sistema no sonarán o reproducirán una copia de lo que esté sonando por los altavoces L o R. Sin embargo, en Ambisonics esto no ocurre, ya que el campo sonoro se descodificará para el sistema de reproducción que deseemos, siendo compatible con cualquier configuración, si bien son necesarios un mínimo de altavoces para una buena resolución espacial. Por tanto una de las ventajas y puntos fuertes de Ambisonics es que **la mezcla y el número de canales Ambisonics es universal e independiente del sistema de reproducción**. Esto quiere decir que mientras un sistema 5.1 común necesita una mezcla propia realizada en dicho formato, haciéndose necesaria una mezcla en estéreo universal y otra en 5.1 a la hora de distribuir material para ambos formatos, en Ambisonics con una mezcla única ésta se puede decodificar y reproducir en cualquier arreglo de altavoces o auriculares, siendo mono y estéreo compatible. Resumiendo, el orden Ambisonics y el número de altavoces de nuestro sistema son independientes.

3.6.1 Binaural

En primer lugar, y como alternativa a los sistemas de reproducción tradicional mediante arreglos de altavoces para sistemas de audios espacial, se va a explicar otro método de monitorización de cualquier mezcla Ambisonics y de audio espacial: la escucha binaural.

La escucha binaural es una técnica de escucha (también puede ser una técnica de grabación) en la que se simula como si el oyente estuviera situado en la misma habitación que la propia banda o lo que estamos reproduciendo, por lo que trata de ser un sistema inmersivo. Para ello hace uso de muchos de los conceptos vistos en 2.1.2. Por tanto es un sistema de reproducción que, mediante procesado digital, pretende recrear los fenómenos acústicos producidos por la escucha humana en un entorno.[21]

Para ello, la tecnología binaural hace uso de fenómenos como la HRTF, vista en 2.1.2, para simular la escucha humana dentro de un espacio sonoro. Es una técnica que parte de la grabación en formato binaural mediante cabezas o sistemas de captación que buscan recrear los mismos fenómenos

⁷Si no tenemos decodificador, escuchando únicamente el primer canal tendremos una versión mono compatible de la escena sonora, y con los dos primeros canales en modo M/S, una versión estéreo compatible.



Figura 3.22:
Sistema de
captación
binaural *Neu-*
mann KU-100.



Figura 3.23: Sistema de capta-
ción binaural *3DIO FreeSpace*.

producidos por el cuerpo humano, tales como la sombra acústica producida por la cabeza o el torso, los filtros producidos por la pinna, etc.

Los sistemas de captación binaural como los expuestos en las figuras 3.22 y 3.23 intentan simular la escucha humana y sus componentes acústicos, tales como sombras, filtros, etc, producidos por la anatomía humana. Los sistemas de reproducción binaural funcionan igual pero aplicados a la parte de *playback*. Es decir, aplican a las señales que alimentan cada transductor los retardos y filtros necesarios para simular dichos fenómenos. Por tanto, la escucha binaural es un tipo de escucha pensada para auriculares. Al colocar nuestros auriculares directamente en nuestra oreja, los fenómenos como la sombra acústica de la cabeza, filtrado de la pinna, ITD, etc, no aparecerán en la señal, y por tanto, se perderá la percepción acústica de estar delante de la fuente sonora grabada. Sin embargo, con la decodificación binaural, estos fenómenos se ven aplicados digitalmente a la señal, y por tanto cuando el oyente escuche por los auriculares, las señales de cada oreja serán diferentes, con sus retardos, variaciones en frecuencia, etc, siendo el resultado mucho más fiel a la realidad. Si además, el oyente dispone de su propia *HRTF* medida, seremos capaces de aplicarla al decodificador, de forma que éste nos dará la respuesta exacta de nuestra cabeza y orejas para el sistema de escucha, mejorando mucho la respuesta espacial de los mismos y el realismo de la grabación.

Sin embargo, la escucha binaural está pensada para ser reproducida mediante auriculares. Esto es, porque al haber alterado cada señal para ser reproducida en binaural, con sus retardos, filtros, etc, éstas están pensadas para ser reproducidas cada una en un oído, sin verse afectadas por el fenómeno del *crosstalk*, el cuál se puede observar en la figura 3.24, y que impide que una grabación o mezcla pensada para binaural se pueda reproducir de manera correcta en un sistema de altavoces. Por tanto, esté fenómeno hace que la escucha binaural esté pensada únicamente para su uso con auriculares, pudiendo mejorar el sistema con la aplicación de matrices de corrección tales como la *HRTF* personal o una ecualización de auriculares para una respuesta más equilibrada.[21]

Si aplicamos esta tecnología al apartado de la reproducción Ambisonics, es una muy buena forma de decodificar nuestras mezclas para, por ejemplo, poder mezclar en formato Ambisonics desde casa y sin disponer de una sala de escucha multialtavoz. Si bien es cierto que hay una diferencia notoria entre ambas, especialmente en la resolución espacial, la escucha mediante decodificación binaural proporciona una respuesta bastante satisfactoria, con una amplitud de campo en la escucha y una resolución espacial bastante sobresalientes para tratarse de una escucha mediante auriculares, y que se ve especialmente incrementada con una *HRTF* personal incluida en el decodificador, lo cuál se verá más adelante en el presente proyecto. Otro de los nichos donde la escucha binaural puede ser una herramienta potente es para la virtualización de sistemas multialtavoz en Ambisonics para su

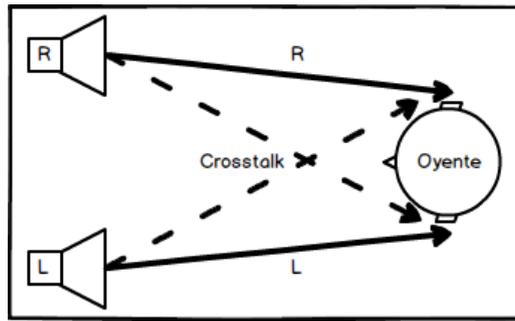


Figura 3.24: Descripción del fenómeno del *Crosstalk*.

monitorización en binaural, como se explica en la sección 4.1.1. Por último, todo el nicho de la VR y los videojuegos han acogido a Ambisonics, y especialmente a la escucha binaural con los brazos abiertos, si bien es cierto que en muchos casos estas tecnologías van implementadas y camufladas dentro de tecnologías y patentes propias que las engloban dentro de otra serie de atributos.

Por tanto, y si pensamos donde podemos reproducir nuestro material Ambisonics de manera externa a DAWs y programas de edición de audio, el reproductor de vídeo multiformato y gratuito *VLC* ofrece la posibilidad de reproducir vídeos con audio en formato Ambisonics de tercer orden, hasta 16 pistas, mediante el códec *spatial audio*, incluido de forma nativa en las últimas versiones de *VLC*. Además, se ofrece la posibilidad de cargar nuestra propia HRTF para una mejora en la escucha, tal y como se observa en la figura 3.25.

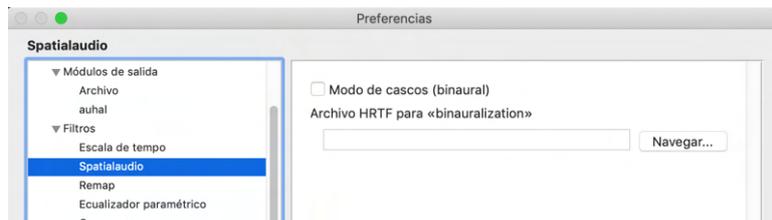


Figura 3.25: Menú de carga para HRTF en el reproductor *VLC*.

Si pensamos en sitios donde podamos subir nuestros vídeos a la red existe *Youtube*, el gigante de los vídeos online nos facilita la posibilidad de subir vídeos con audio Ambisonics de primer orden además de primer orden más estéreo. Además, nos ofrece el VST para monitorizar nuestro audio antes de subirlo y poder tener una referencia del resultado a la hora de subirlo a la plataforma. Dicho VST se conoce como *Resonance Audio Monitor* (RAM), y está disponible de manera gratuita para gente que desee subir sus vídeos con audio espacial a *Youtube*. RAM no posee ningún tipo de control, ya que únicamente se inserta en el bus que deseemos utilizar como bus de salida y automáticamente realiza una decodificación a escucha binaural sin que el usuario pueda ni tenga que configurar nada, tal y como se puede observar en la figura 3.26.[20]

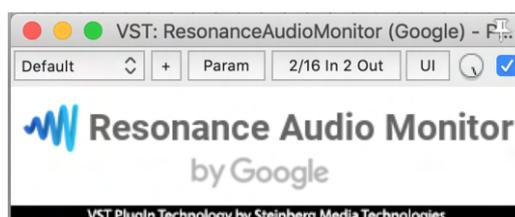


Figura 3.26: Decodificador binaural de Google: *Resonance Audio monitor*.

Por último, y donde más opciones tenemos a la hora de decodificar audio binaural desde Ambisonics es a través de un DAW, y especialmente en este proyecto se recomienda como ya se viene haciendo a lo largo de todo el proyecto, *Reaper*. Se recomienda *Reaper* por ser un DAW con una licencia de prueba completa y sin limitación de tiempo, un precio de compra muy competitivo, además de por poseer una serie de características que lo hacen ideal para trabajar con Ambisonics, tales como buses de hasta 64 canales, o una matriz de *routing* muy completa y adecuada a la hora de configurar salidas multialtavoz por ejemplo. Ya en *Reaper*, las opciones que tenemos para monitorización binaural son múltiples, debido principalmente a que muchos fabricantes incluyen un decodificador binaural de Ambisonics en su suite de *plugins*. En la figura 4.9 podemos observar el decodificador a binaural que se va a utilizar a lo largo del proyecto. Se ha elegido especialmente por algunas características que lo hacen único, como la posibilidad de sincronizarlo mediante el protocolo OSC con un *headtracker* externo, tal y como se puede ver en la sección 4.1.2, insertar nuestra propia HRTF o corrección de auriculares mediante archivo estándar SOFA, para cargar respuestas al impulso o la versatilidad que ofrece a la hora de seleccionar el orden en el que queremos trabajar o la codificación de canales que vamos a usar. Por todo ello, el *AmbiBIN* es uno de los decodificadores Ambisonics - binaural más completos, y además, su uso es gratuito y multiplataforma mediante formato VST.

3.6.2 *Playback* en arreglos de altavoces

La otra forma disponible para trabajar con Ambisonics y poder monitorizar el campo sonoro es a través de un array de altavoces correctamente distribuido, ajustado y calibrado. Es un método más costoso, tanto en recursos como en dificultad de instalación que la monitorización binaural, si bien los resultados para un sistema correctamente instalado son bastante sorprendentes, además de las posibilidades que ofrece para laboratorios, salas de exposiciones audiovisuales, cines y hasta conciertos y teatros. Antes de pasar a explicar los tipos de modelos que podemos diseñar, así como los tipos de decodificadores que podemos utilizar, se va a ver una serie de factores o limitaciones derivados del uso de múltiples altavoces, lo cuál puede provocar problemas de fase, resolución espacial, etc, si éstos no están correctamente ajustados.[22]

- **Efecto *Near-field* o de campo cercano.**- En su teoría, Ambisonics asumía frentes de onda planos, lo que para un sistema de altavoces, éstos deberían estar situados en el infinito. Los altavoces son fuentes puntuales, lo que provoca frentes de onda esféricos si el oyente no se sitúa en una posición cercana al infinito para la longitud de onda más grande. Esto provoca en la escucha un aumento en baja frecuencia similar al efecto de proximidad que se da en los micrófonos direccionales. Por tanto, y como método de control de este efecto, una compensación del campo cercano mediante una ecualización en la zona de baja frecuencia puede ser muy beneficiosa para la escucha.
- **Distancia de los altavoces y angulación.**- Bajo la misma suposición de un frente de ondas plano, éste nos permite modificar y alterar la distancia de los altavoces dentro del array bajo unos límites razonables sin hacer que el decodificador llegue a comportarse de manera errática, únicamente ajustando la distancia extra con los retardos y el ajuste de ganancia pertinente para que en el centro de la esfera, todos los altavoces lleguen en fase y con la misma potencia SPL, además de la pertinente compensación de campo cercano en cada altavoz. Por tanto, la distancia a la que los altavoces estén situados del centro no es pertinente siempre que se tenga en cuenta, se ajuste y se corrija de manera satisfactoria.

Todo esto nos da a entender que la distancia de los altavoces al centro de la esfera es el mayor grado de libertad a la hora de diseñar ajustes idealizados dentro de salas reales. Uno de los factores que pueden limitar esta condición es la reverberación de la sala, la cual puede provocar

relaciones desiguales para altavoces cuya distancia al centro varíe significativamente. Otro de los factores limitantes es la gestión de la potencia por parte de los altavoces más alejados, la cuál puede presentar limitaciones en la dinámica debido a un aumento excesivo de la ganancia del amplificador para compensar la mayor distancia al centro. Por tanto, si respetamos correctamente al ángulo de colocación de los altavoces y la distancia de éstos al centro es bastante versátil siempre y cuando se realice una correcta compensación y ajuste del equipo, además de que éste ofrezca una potencia suficiente para funcionar en un régimen de trabajo adecuado.

- **2D vs 3D.-** Si bien es cierto que uno de los grandes atractivos de Ambisonics es la posibilidad de decodificar y reproducir el audio en 360° , también lo es que los sistemas de reproducción 3D con altavoces en altura ofrecen una localización espacial más inestable en altas frecuencias frente a los sistemas únicamente horizontales. Por ello, y para sistemas que requieran de alguna escucha ocasional únicamente horizontal, se recomienda un anillo horizontal denso con un mayor número de altavoces que los anillos en altura.
- **Problemas de fase.-** En un sistema de varios altavoces sonando a la vez, donde en muchos casos la señal emitida por varios de ellos contendrá contenido correlacionado, si el oyente se encuentra en movimiento puede llegar a experimentar una especie de efecto *Doppler*, o alteración del timbre percibido, lo que puede llevar a una localización mucho más pobre. Además, este tipo de artefactos sonoros en la fase suelen verse incrementados en salas secas y con sistemas muy bien calibrados. Para reducir dicho efecto, se pueden incorporar altavoces en altura para suavizar el efecto. Es un efecto que se da sobre todo en sistemas donde el oyente camina o está en movimiento, mientras que para sistemas donde el oyente está sentado, el efecto es mucho menor, a no ser que los patrones de interferencia sean tan densos que el oyente pueda llegar a percibir estos artefactos únicamente con los movimientos de la cabeza.
- **Nº de altavoces vs orden Ambisonics.-** Tal y como afirma Solvang[23] en sus investigaciones, el aumento del número de altavoces por encima del mínimo no garantiza mejores resultados, pudiendo llegar a ser incluso perjudicial para el resultado final. El porqué es fácil de imaginar: a mayor número de altavoces más resolución angular, menos diferencia de señal entre altavoces colindantes, mayor correlación entre señales, mayor diafonía y por tanto, un filtro peine cada vez más grande en la zona de escucha, especialmente para oyentes en movimiento. Por tanto, es interesante poder diseñar sistemas de altavoces modulares donde poder añadir o quitar altavoces dependiendo del orden Ambisonics que vayamos a reproducir y sin que esto afecte a la posición de los mismos. Como ejemplo podemos tener un octógono de tercer orden, el cual encierra en sus vértices la posibilidad de un cuadrado de primer orden. Con un sistema como este, y en caso de querer reproducir material de primer orden, es preferible decodificar en un cuadrado regular que en un octógono de tercer orden, ya que esto provocaría el problema indicado de la diafonía y el filtro peine.

Por otro lado, y según Zotter y Frank[24], la decodificación de Ambisonics se ha tratado en profundidad en el pasado, especialmente para los resultados que ofrecen los sistemas de primer orden, donde la localización espacial no es estable a lo largo de toda la esfera, y por otro lado por la obligación de sistemas regulares y uniformes para una correcta decodificación en ordenes superiores. Según Zotter y Frank la solución más práctica a la hora de decodificar para sistemas de altavoces en Ambisonics es el uso de la decodificación “ambisónica” total, conocida como **AIRAD**, especialmente por las posibilidades que ofrece a la hora de situar altavoces imaginarios en nuestro sistema así como la posibilidad de realizar arreglos irregulares sin que por ello tengamos que pagar un alto precio en la calidad de la decodificación. Por otro lado, este sistema no posee restricciones a la hora de elegir el orden Ambisonics que deseemos.

Por tanto, el número de direcciones reproducibles θ_l o φ_l suele ser finito y discreto, y viene

representado por las direcciones de los altavoces colindantes. Es por ello que la distribución direccional de los altavoces colindantes no es un tipo de diseño en t^8 , cuyo patrón viene definido por $t \geq 2N + 1$, ni tampoco es un diseño de polígonos regulares, con $L \geq 2N + 2$ altavoces para sistemas 2D.

Tras haber visto algunos de los factores que debemos tener en cuenta a la hora de diseñar arreglos de altavoces para decodificar Ambisonics, y a pesar de que se ha comentado que el decodificador tipo AllRAD es una solución completa y versátil aún con sus limitaciones, se van a ver los tipos de decodificadores que existen actualmente en el mercado, la base teórica en la que se apoya su funcionamiento, y cuáles son sus características principales, ventajas, desventajas, etc:

Sampling Ambisonics Decoder (SAD)[24]

Se trata del tipo de decodificación más sencillo. Para un sistema de dos dimensiones ($D = 2$) y tres ($D = 3$), se utiliza la matriz $Y_N = [Y_N(\theta_1), \dots, Y_N(\theta_L)]$ que contiene los respectivos armónicos circulares o esféricos $Y_N(\theta)$ muestreados o *sampleados* en la dirección de los altavoces $\{\theta_L\}$,

$$D = \sqrt{\frac{S_D - 1}{L}} Y_N^T,$$

con la circunferencia del círculo unidad denominada $S_1 = 2\pi$ o la superficie de la esfera unidad escrita como $S_2 = 4\pi$. El factor $\sqrt{\frac{S_D - 1}{L}}$ expresa que cada altavoz sintetiza una fracción de la medida de E en el círculo o esfera de las direcciones *surround*. No obstante, un decodificador basado en muestreo no producirá nunca mediciones de volumen y apertura espacial perfectamente constantes, E , $\|\mathbf{r}E\|$, ni tampoco una orientación correcta para la medida de localización $\mathbf{r}E$ si la disposición de los altavoces no es óptima.

Como ejemplo, en lo que se refiere a sonoridad, para un desplazamiento hacia regiones o zonas donde los altavoces ofrecen una cobertura direccional deficiente, se aprecia como el muestreo omite el lóbulo principal para la función de desplazamiento, dando lugar a una sonoridad en esa región notablemente reducida.

Mode Matching Decoder (MMD)[24]

El método *Mode-Matching* produce un diseño de decodificador diferente en sus fundamentos de funcionamiento al SAD. Su concepto se basa en recodificar el vector de ganancia g de los altavoces para cualquier dirección de *panneo* o panoramización θ_S mediante la matriz de codificación $Y_N = [Y_N(\theta_1), \dots, Y_N(\theta_L)]$ para todas las direcciones de los altavoces $\{\theta_l\}$. Lo ideal por tanto es que el resultado recodificado coincida con la codificación de la dirección de *panneo* con los lóbulos laterales suprimidos

$$\mathbf{Y}_N \mathbf{g} = \text{diag}\{\mathbf{a}_N\} \mathbf{y}_N(\theta_S),$$

Utilizando la expresión $g = \mathbf{D} \text{diag}\{\mathbf{a}_N\} \mathbf{y}_N(\theta_S)$ para las ganancias de paneo obtenemos que

⁸Un diseño en t para una esfera viene definido por una distribución uniforme de los nodos, en nuestro caso altavoces, a lo largo de dicha esfera.

$$\mathbf{Y}_N \mathbf{D} \text{diag}\{\mathbf{a}_N\} \mathbf{y}_N(\theta_S) = \text{diag}\{\mathbf{a}_N\} \mathbf{y}_N(\theta_S),$$

$$\Rightarrow \mathbf{D} = \sqrt{\frac{L}{S_{D-1}}} \mathbf{Y}_N^T (\mathbf{Y}_N \mathbf{Y}_N^T)^{-1}$$

por lo que se observa que el decodificador \mathbf{D} debe estar invertido a la matriz \mathbf{Y}_N , es decir, $\mathbf{Y}_N \mathbf{D} = \mathbf{Y}_N \mathbf{Y}_N^T (\mathbf{Y}_N \mathbf{Y}_N^T)^{-1} = \mathbf{I}$. Para que exista la matriz inversa de $\mathbf{Y}_N \mathbf{Y}_N^T$, es necesario tener como mínimo el mismo número de altavoces que de armónicos para el orden en el que trabajemos, es decir, $L \geq (N+1)^2$ para $D = 3$ o $L \geq 2N$ para $D = 2$. Sin embargo, este criterio no es suficiente, ya que cuando nos encontremos en direcciones con cobertura pobre, la inversión aumentará el volumen, de modo que el resultado podrá ser a menudo numéricamente mal acondicionado para $(\mathbf{Y}_N \mathbf{Y}_N^T)^{-1}$, a menos que se diseñe una plantilla de altavoces uniforme, como poco. El método *Mode-Matching*, por tanto está mal acondicionado para diseños de altavoces semicirculares o semiesféricos.

Energy Preserving Ambisonics Decoder (EPAD)[24]

Si tratamos de establecer una sonoridad invariable en toda la panorámica del campo sonoro para decodificar en diseños de altavoces que sean no uniformes, se puede asegurar una medida de sonoridad E constante si aplicamos que $\mathbf{D}^T \mathbf{D} = \mathbf{I}$, lo cuál de otro modo sólo se puede conseguir en diseños del tipo $t \geq 2N$.

Por tanto, podemos buscar una matriz de decodificación \mathbf{D} cuyas entradas estén más cerca del decodificador de muestreo bajo la restricción de ser ortogonales por columnas.

$$\|\mathbf{D} - \sqrt{\frac{S_D - 1}{L}} \mathbf{Y}_N^T\|_{Fro}^2 \rightarrow \min$$

sujeto a $\mathbf{D}^T \mathbf{D} = \mathbf{I}$

La descomposición del valor singular

$$\mathbf{Y}_N^T = \mathbf{U}[\text{diag}\{S\}, \mathbf{O}]^T \mathbf{V}^T$$

que podemos usar para crear

$$\mathbf{D} = \mathbf{U}[\mathbf{I}, \mathbf{O}]^T \mathbf{V}^T$$

únicamente reemplazando los valores singulares S por unos. Este decodificador es ortogonal en forma de columna, ya que la descomposición en valores singulares proporciona $\mathbf{U}^T \mathbf{U} = \mathbf{I}$ y $\mathbf{V} \mathbf{V}^T = \mathbf{I}$, y como consecuencia $\mathbf{D}^T \mathbf{D} = \mathbf{I}$. Para llegar hasta ahí, la secuencia es la siguiente:

$$\mathbf{D}^T \mathbf{D} = \mathbf{V}[\mathbf{I}, \mathbf{O}] \mathbf{U}^T \mathbf{U}[\mathbf{I}, \mathbf{O}]^T \mathbf{V}^T = \mathbf{V}[\mathbf{I}, \mathbf{O}] \mathbf{I} [\mathbf{I}, \mathbf{O}]^T \mathbf{V}^T = \mathbf{V} \mathbf{V}^T = \mathbf{I}$$

Por tanto, el decodificador para preservar la energía en esta versión básica requiere de $L \geq 2N + 1$ altavoces en sistemas 2D o $L \geq (N + 1)^2$ en 3D para funcionar.

NOTA: Se puede observar que si las direcciones de los altavoces funcionan como un diseño $t \geq 2N$, los métodos de decodificación SAD, MMD y EPAD son equivalentes.

All-Round Ambisonics Decoding (AIIRAD)[24]

Llegados a este punto es importante destacar que los decodificadores aquí citados no son los únicos disponibles, si bien son los más efectivos y disponibles en las aplicaciones habituales en el mercado. Por ejemplo, y antes de hablar más en profundidad del decodificador **AIIRAD**, en la figura 3.27 observamos como la suite SPARTA de la universidad de AALTO en Finlandia, nos permite el poder trabajar en dos bandas de frecuencia diferentes con una frecuencia de corte seleccionable por el usuario, y aplicar un tipo de decodificación diferente en cada una de las bandas, con la posibilidad de elegir decodificación de tipo SAD, MMD, EPAD y AIIRAD. Esto es muy interesante de cara a separar la banda de frecuencias grave de las medias y agudas, de forma que podamos aplicar distinto tipo de decodificador dependiendo de la situación, pudiendo controlar mejor la zona problemática de bajas frecuencias.

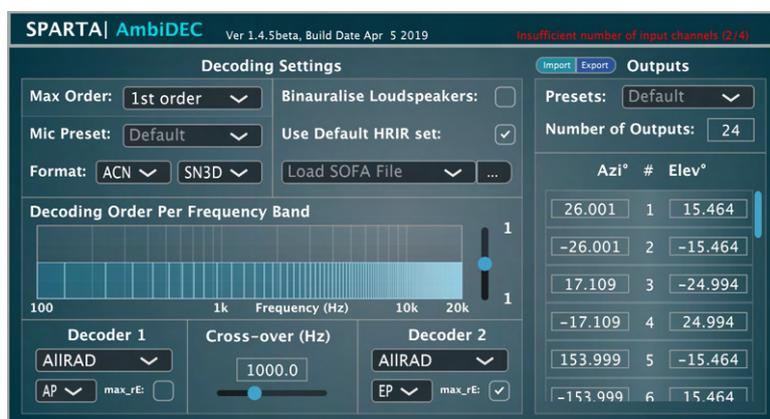


Figura 3.27: *AmbiDEC* de la suite SPARTA con decodificación *AIIRAD* para ambas bandas de trabajo.

La decodificación completa en Ambisonics, conocida como **AIIRAD**, basada en VBAP[27], se ha mostrado como una de las más prácticas y efectivas a la hora de decodificar contenido Ambisonics. En parte por su posibilidad de insertar altavoces imaginarios dentro de nuestro arreglo de altavoces, lo cual hace del decodificador una herramienta muy versátil. Otra de sus características es la de no tener un orden en el que su decodificación es óptima, ya que para cualquier orden es posible trabajar con un decodificador de este tipo. Esta posibilidad de trabajar en varios órdenes facilita el control de las fluctuaciones de energía en ciertas zonas del campo sonoro, tal y como otro tipo de decodificadores sí ofrecen para órdenes bajos, así como el control del panning y la direccionalidad de las fuentes dentro del campo sonoro.

La demostración matemática del funcionamiento de los decodificadores tipo **AIIRAD** es demasiado compleja para ser demostrada en el presente proyecto al carecer de interés para el mismo. Si se desea consultar, se pueden consultar los trabajos de *Zotter y Frank*[25] y los de *Batke y Keiler* [26].

Por tanto, la decodificación tipo **AIIRAD** se ajusta al propio concepto de Ambisonics para arreglos de altavoces irregulares. La decodificación Ambisonics convencional únicamente es sencilla y efectiva si se dispone de una distribución óptima de los altavoces, para cuyo caso consigue una distribución de la energía independiente de la dirección y una dispersión de la misma, es decir, un volumen y una anchura de las fuentes imaginarias adecuados. Todo ello lo convierte en un tipo de decodificador completo pero sobre todo versátil, y con una tolerancia a los arreglos de altavoces no uniformes muy buena. Este detalle podría pasar desapercibido, pero es importante ya que en la mayoría de los casos, para salas físicas reales será muy complicado hacer una distribución lineal y uniforme de los altavoces a lo largo de la misma, lo que convierte a este tipo de decodificadores en una poderosa herramienta para la decodificación en Ambisonics.

Decodificación 3D en diseños hemisféricos

Tal y como ya se ha comentado, por norma general los arreglos óptimos se dan para esferas uniformes con arreglos tipo T, donde cada altavoz tiene su opuesto a lo largo de la esfera, y además con una cobertura uniforme y lineal a lo largo de toda la superficie. Para casos de investigación puede ser viable un arreglo de este tipo, pero para grandes audiencias esto se vuelve complicado. Para este tipo de reproducciones con público se torna mucho más sencillo y práctico un sistema hemisférico, donde no sea necesario colocar altavoces a nivel de suelo, colocando únicamente un anillo a la altura de escucha de los oyentes, y una serie de altavoces en altura para la información sonora superior.

Este tipo de casos se puede dar por ejemplo para salas grandes con un suelo sólido, donde se vuelve complicado la colocación de altavoces. Pero tal y como se ha dicho, esta situación hace que no se permita la decodificación por muestreo, (SAD), con diseños óptimos en T que cubran todas las direcciones. Sin embargo, y como también se ha comentado, los métodos de decodificación EPAD y AllRAD ofrecen una versatilidad extra para la decodificación en arreglos que no sean de tipo T, y, sin embargo, todavía requieren de cierta precaución y cuidado cuando se utilizan con diseños y arreglos de altavoces hemisféricos.

- **AllRAD con diseños hemisféricos.**

Debido al uso de VBAP⁹, la decodificación de tipo AllRAD es, actualmente y de manera comparativa, la más robusta para arreglos y diseños de altavoces irregulares. Sin embargo, para un tipo de arreglo hemisférico o de media esfera, dicha disposición no contiene ningún vector de dirección de ningún altavoz que apunte a la zona inferior de la esfera, de manera que se podría omitir dicha información en el material reproducido. Esto, a pesar de parecer algo lógico, nos induciría a error, ya que la función de panoramización Ambisonics implica una dispersión direccional, de modo que cuando paneamos un sonido exactamente al horizonte, éste también produce contenido por debajo de éste, cuya omisión causa:

- Una reducción en el volumen emitido.
- Percepción de una ligera elevación de la fuente por encima de la línea del horizonte.

Para poder evitar este tipo de comportamiento y que aparezcan las carencias mencionadas durante la reproducción, la inserción de altavoces imaginarios se hace necesaria y fija este comportamiento indeseado. Para el diseño de arreglos de altavoces hemisféricos, no es necesario mezclar la señal del altavoz imaginario en el nadir de la esfera para estabilizar tanto el volumen como la localización de fuentes en el horizonte o cercanas a él. Para ello, las contribuciones de señal por debajo, pero cercanas al horizonte ayudan en gran medida a que los altavoces horizontales estabilicen tanto su volumen como su localización, lo que hace seguro disponer de la señal que alimentaría a un altavoz imaginario en el nadir de la esfera sin pérdida de volumen. De esta manera, la localización de los altavoces horizontales se vuelve a bajar a su posición y no creará imágenes fantasma y artefactos en la localización espacial de las fuentes.

Como ejemplo tenemos las figuras 3.28 y 3.29. En ellas se ha diseñado un arreglo estándar de 8 altavoces en el anillo central y 4 superiores, haciendo así un arreglo hemisférico. Para comprobar el efecto mencionado en esta sección, se ha situado un altavoz en la posición de nadir de la esfera, y se ha calculado su mapa de fluctuación de energía y el parámetro $across-rE$ para la anchura de fuente tanto para el altavoz de nadir imaginario, como real.

Al contrario de lo que cabría esperar, y tal y como se ha mencionado, los resultados arrojados nos dan un mapa de fluctuación de energía más uniforme y denso para un altavoz de nadir imaginario,

⁹VBAP.- *Vector Base Amplitude Panning.*

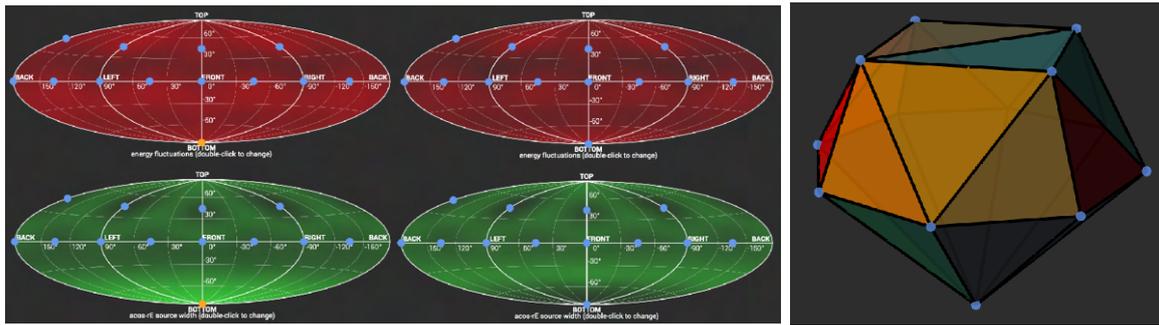


Figura 3.28: *Izda.* Altavoz en posición nadir imaginario. *Dcha.* Arreglo hemisférico 8+4 en 3D. Altavoz en posición nadir real.

así como unos niveles más estables en la zona central de la esfera, lo que conoceríamos como “horizonte” de la misma.

3.7 Distribución Ambisonics.

La manera en la que podemos distribuir nuestro material Ambisonics en ocasiones es confusa, ya que dependiendo de nuestro propósito o del *target* al que vaya dirigido, será conveniente elegir uno u otro método o formato. Dependiendo del punto de la cadena de trabajo expuesta en la figura 3.2 en el que nos encontremos, el producto final será diferente. Principalmente esta diferencia va a venir marcada por el tipo de audio que queramos entregar al cliente, precodificado o no, predecodificado o no, etc.

Todo esto se debe principalmente al hecho de que una grabación Ambisonics puede venir representada de varias maneras o formatos diferentes mediante un conjunto de señales. En general, dicho conjunto de señales puede transformarse de uno a otro, teóricamente sin pérdidas, si bien algunas transformaciones son más sencillas de llevar a cabo que otras, así como algunas logran una mayor precisión frente al material original que otras.[28]

- **A-Format.**- Este tipo de formato de trabajo para las señales Ambisonics ya se ha visto por encima, y es el correspondiente a las señales procedentes de un micrófono Ambisonics. Al principio únicamente se utilizaba para designar las señales procedentes de un micrófono de primer orden, es decir, a las cuatro señales de las cuatro cápsulas cardioides de un micrófono tetraédrico de campo sonoro. Ahora bien, dependiendo del micrófono, el fabricante, etc, las características de estas cápsulas pueden variar entre distintos diseños de micrófono, así como el significado exacto de las señales de formato A. Esto es debido a que cada sistema de micrófono utiliza un procedimiento para realizar la conversión de A a B, algunos llevan conversión por hardware, otros entregan las señales en formato A y realizan la conversión vía software desde el PC, etc.

Por tanto, aunque el formato A se entiende, en general, como la señal procedente de un micrófono tetraédrico, es decir, de primer orden, hay cierto consenso en que este formato designa más bien las señales físicas antes de la conversión a formato B, y por tanto también abarcará las señales sin convertir de un micrófono HOA.

Como curiosidad, se ha demostrado que algunas tareas matemáticas utilizadas para el procesamiento de señales Ambisonics se dan mejor en formato A que en formato B; la reverberación por ejemplo se ha utilizado de esta manera. Para este caso, se generan un conjunto de señales estándar en formato A a partir del formato B, y de esta manera aplican la reverberación en

formato A, antes de que estas señales se corrijan mediante los filtros explicados en la conversión A a B, en la sección 3.3.

- **B-Format.**- Se trata del formato básico y universal que se utiliza para el almacenamiento y manipulación de audio en formato Ambisonics. Este formato contiene y consiste básicamente en los armónicos esféricos que definen el campo sonoro, siendo el número de éstos mayor dependiendo del orden en el que estemos trabajando. Tal y como ya se ha visto, un Ambisonics en formato B de primer orden viene definido por una señal correspondiente al orden 0, la cuál es conocida como W y representa una señal mono omnidireccional, y las tres señales correspondientes al 1^{er} orden, conocidas como X, Y y Z, las cuáles corresponden a tres micrófonos ortogonales de figura de ocho colocados todos en el punto central de la esfera del campo sonoro.

Ninguna de estas señales es utilizada para alimentar ningún altavoz directamente, si no que todas ellas definen el campo sonoro completo y permiten la manipulación del mismo, así como la posibilidad de decodificar el campo sonoro para generar las señales que sí alimentarían los altavoces, girar el campo sonoro, y otra serie de transformaciones sobre el mismo. Además, al tratarse de las señales que definen el campo sonoro antes de ninguna conversión ni decodificación, hacen del formato B el elegido para el almacenamiento y transmisión de material Ambisonics, siempre que lo que se desea entregar sea el campo sonoro completo para que el usuario final pueda decidir qué desea hacer con el.

Cuando nos adentramos en los formatos HOA, se necesitan armónicos adicionales para la definición del campo sonoro. Por ejemplo, para segundo y tercer orden, se utilizaron letras adicionales, mientras que existe el estándar ACN (*Ambisonics Channel Number*), el cual utiliza números y no tiene un límite de orden superior. Para dichos órdenes superiores, existe un problema debido a la flexibilidad de la nomenclatura. En la comunidad Ambisonics existe cierto debate sobre cómo y en qué orden deben ser listadas las señales de orden superior. Además, también existe cierta controversia sobre si el formato B debe limitarse al conjunto de cuatro señales de primer orden o debe utilizarse indistintamente para cualquier orden Ambisonics.

- **C-Format.**- Se trata de un formato pensado para distribución al consumidor final. En los inicios de la tecnología Ambisonics no era práctico considerar la distribución de las 4 señales en formato B. Por tanto, se diseñó un formato que representase esas señales de forma alternativa, la cual se llamó *UHJ*. El *UHJ* proporcionaba dos canales (L y R), compatibles con el estéreo, y cuya suma era un señal mono compatible. L y R fueron diseñadas de tal forma que se podía generar un sonido envolvente horizontal completo a partir de esas dos señales, si bien éste tenía una resolución reducida. Además, una tercera señal, conocida como T, podía combinarse con L y R para ampliar el campo y regenerar las señales originales W, X e Y, y además, una cuarta, llamada Q, que transportaba la altura de la Z original.

Llegados a este punto, y con el *UHJ* definido, se hicieron propuestas comerciales para transmitir varios tipos de combinaciones de este conjunto de señales LRTQ en radio FM, pero dichas propuestas no fueron aceptadas por la industria, lo que hizo que el formato *UHJ* completo muriese. Sin embargo, el sistema de L y R para representar una escena surround horizontalmente (y conocida universalmente como *UHJ Stereo*, si bien originalmente se denominó *BHJ*) fue acogido por la discográfica *Nimbus* con entusiasmo, lo que generó un catálogo completo grabado en Ambisonics y publicado en formato *UHJ Stereo*. Existen otras discográficas con material editado de la misma forma, si bien éstas fueron contadas excepciones.

- **D-Format.**- Inicialmente se utilizó como nomenclatura para un conjunto de señales Ambisonics que habían sido decodificadas como alimentación para una serie particular de altavoces. A estas conjunto de señales correspondientes a cada altavoz se le denominó formato D. Sin embargo, el termino actualmente está en desuso (ver *G-Format*, bajo).

- **E-Format.**- Si pensamos en el formato C (*UHQ*), la primera parte de la decodificación puede ser recuperar algo parecido al formato B. Sin embargo, y si pensamos en el caso horizontal, si la señal T no está, o no posee el ancho de banda completo, la reconstrucción será incompleta e imperfecta. *Michael Gerzon* se refirió a las señales recuperadas desde el formato C pero incompletas como formato E, si bien no se tiene constancia de que ésta sea una nomenclatura habitual.
- **G-Format.**- Llegados al punto en el que transmitir cuatro o más canales se hizo práctico, la forma lógica de transmitir campos sonoros en Ambisonics fue el formato B. Sin embargo, en ese punto, era viable transmitir el formato B completo, pero el tema de los decodificadores era algo complicado, sin opciones software disponibles por la baja capacidad de procesamiento de los ordenadores de la época. Es por ello que *Geoffrey Barton*[29] propuso en aquella época que, en lugar de distribuir el material en formato B y que el usuario tuviera problemas para su decodificación, lo mejor era que los esfuerzos de la industria se concentraran en la distribución de fuentes de altavoces para los diseños más populares, como el 5.1. *Gerzon* se refirió a esto como formato *Geoffrey*, quedando como formato G. En la práctica, el término *G-format* se utiliza en lugar de *D-format* para cualquier conjunto de fuentes de altavoces.

Como curiosidad, la discográfica Nimbus ha decodificado algunas de sus grabaciones y las ha comercializado en formato DVD-A para poder ser reproducidas en sistemas cuadrafónicos; sin embargo, el DVD-A nunca terminó de despegar como formato comercial, lo que ha limitado en exceso su atractivo e impacto.

3.7.1 Formatos de archivos Ambisonics. Metadatos.

Cuando se trata de la transmisión y el intercambio de archivos Ambisonics, son varios los formatos de archivo disponible para contener los múltiples canales del formato B. Se habla de formato B por ser el único de los formatos explicados que realmente representa un campo sonoro 3D en Ambisonics, ya que el resto son únicamente formatos multicanal para alimentar un sistema de altavoces cualquiera, el formato A que todavía no representa un campo Ambisonics con sus armónicos esféricos, etc.

Por todo ello, y para pistas de formato B, se hizo necesario un tipo de archivo para el almacenamiento y transmisión de esas pistas multicanal, para lo cual se definió un formato propio con una serie de propiedades adicionales, como una serie de metadatos encargados de llevar la información del tipo de normalización u orden de canales, duración, etc, así como el hecho de ser un archivo de audio multicanal, con lo que ello supuso hace unos años.

- **.AMB.**- En el año 2001, Richard Dobson [30] propuso y definió un nuevo formato de archivo basado en la propuesta *WAVE_FORMAT_EXTENSIBLE* de *Microsoft*, basada en el formato de archivo de audio WAV. Estaba basado en la codificación FuMa.

Viniendo del formato WAV, hereda de éste un tamaño máximo por archivo de hasta 4GB, lo que ya da a entender la antigüedad y limitación de este formato, especialmente si nos proponemos trabajar con pasajes de una longitud considerable o de órdenes superiores.

Los archivos de tipo *.amb* posee un sufijo propio y un subtipo de archivo llamado *Globally Unique Identifier* en sus datos de cabecera para poder identificarse de otros contenidos multicanal. En dicho identificador, los valores pueden ser los siguiente:

```
SUBTYPE_AMBISONIC_B_FORMAT_PCM {000000001-0721-11d3-8644-C8C1CA000000}
SUBTYPE_AMBISONIC_B_FORMAT_IEEE_FLOAT {000000003-0721-11d3-8644-C8C1CA000000}
```

El primero de ellos corresponde a muestras enteras, mientras que el segundo se refiere a muestras en coma flotante.

La definición del archivo recomienda que cada archivo de audio posea un trozo denominado como *PEAK*, el cual contendrá la posición y el valor de la muestra más alta de cada uno de los canales. Por otro lado, los canales dentro de un archivo *.amb* están intercalados y se omite cualquier canal que no se esté utilizando, reduciendo así el ancho de banda. De esta manera, el formato permite identificar archivos de audio de órdenes mixtos o híbridos mediante el número de canales, todo ello según la siguiente tabla:

Nº canales	Orden		Descripción	Plantilla de componentes
	Horiz.	Vert.		
1	0	0	Mono	W
2	1/2	0	Estéreo M/S	WY
3	1	0	FOA horizontal	WXY
4	1	1	FOA 3D	WXYZ
5	2	0	SOA horizontal	WXY,UV
6	2	1	SOA horizontal, FOA 3D	WXYZ,UV
7	3	0	TOA horizontal	WXY,UV,PQ
8	3	1	TOA horizontal, FOA 3D	WXYZ,UV,PQ
9	2	2	SOA 3D	WXYZ,RSTUV
11	3	2	TOA horizontal, SOA 3D	WXYZ,RSTUV,PQ
16	3	3	TOA 3D	WXYZ,RSTUV,KLMNOPQ

En la biblioteca de C denominada *libsndfile*, libre y de código abierto, se incluye soporte para el formato *.amb* desde el año 2007.

Si bien actualmente es un formato que tiende al desuso y el abandono, el formato creado por *Dobson* ha sido fundamental en la historia de Ambisonics para que el contenido nativo se haya podido compartir y tratar de manera accesible por la comunidad de entusiastas, así como para facilitar el camino a la investigación y el despliegue de Ambisonics de manera más o menos masiva o comercial. Si bien es un formato con un orden máximo de trabajo de tercer orden, sus capacidades son, y sobre todo han sido, suficientes para el crecimiento de Ambisonics.

- **AmbiX.**- Con la llegada de los nuevos sistemas HOA se hizo necesario un nuevo estándar de formato de archivo para la codificación y transmisión de material Ambisonics, capaz de evitar las limitaciones del formato **.amb** tales como el tamaño máximo de archivo o el orden máximo Ambisonics. Para ello en el *Ambisonics Symposium 2011* se propuso el formato AmbiX [31] como nuevo estándar.

AmbiX deriva y adopta el formato de audio *Core Audio Format* de Apple, también conocido como *.caf*, y permite cualquier orden Ambisonics de trabajo, por alto que sea, así como cualquier tamaño de archivo sin límite superior. El formato *.caf* funciona como archivo contenedor, siendo el tipo de dato de tipo PCM lineal con la posibilidad de trabajar en tamaños de palabra de 16, 24 o 32 bits en punto fijo, o 32 bits en coma flotante, así como cualquier frecuencia de muestreo que nos permita el formato *.caf* [32]. Como ya se ha comentado, AmbiX utiliza un orden de canales ACN con una normalización SN3D.

El formato AmbiX básico exige un conjunto completo de señales que representen la esfera completa. Además, su orden puede deducirse de forma única a partir del número de canales. Se trata de un formato cuya información única viene dada por la cabecera mínima requerida por el formato *.caf*, excluyendo cualquier otro tipo de metadato en el archivo.

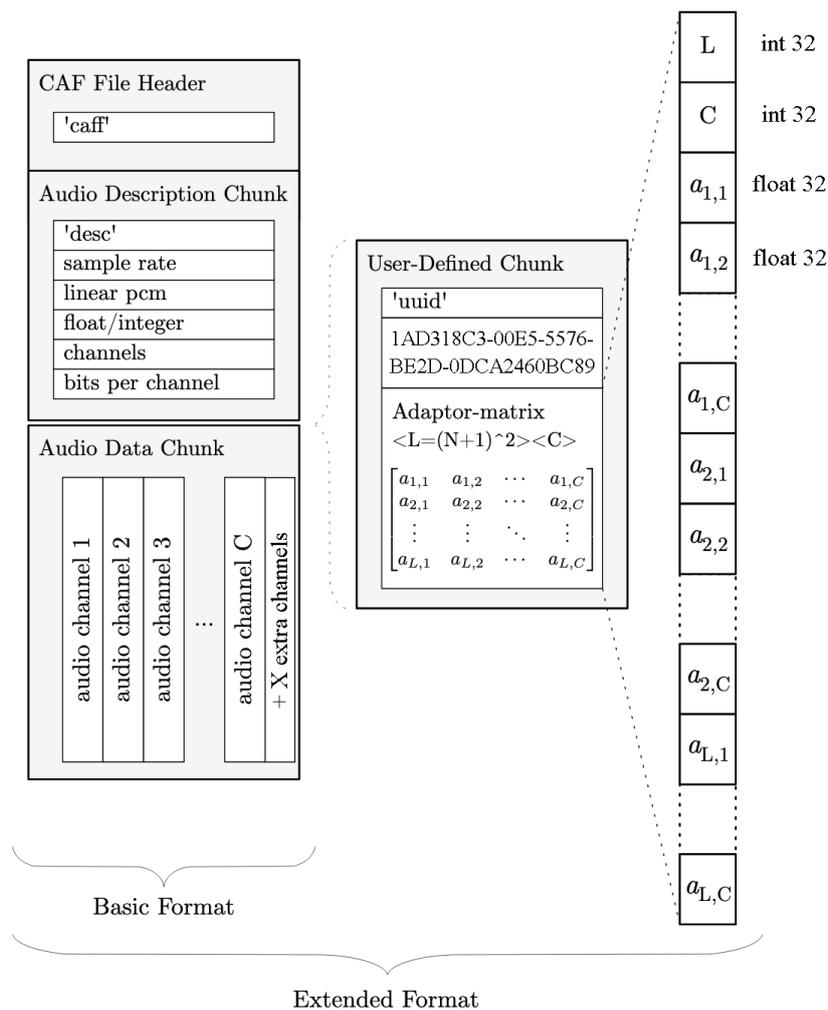


Figura 3.30: Estructura de archivo para los formatos AmbiX básico y extendido. [31]

Además, para escenas sonoras que no son completamente perifónicas o 3D, AmbiX ofrece un formato *Ambix extended*, el cual posee una matriz de adaptación en la *User Defined Chunk*, o parte definida por el usuario, dentro de la cabecera, donde se incluye el UUID (*Universally User Identifier*):

1AD318C3-00E5-5576-BE2D-0DCA2460BC89

Si este identificador está presente, las señales no se reproducirán directamente si no que pasarán por la matriz de corrección incluida en el propio archivo, de manera que el campo sonoro se adaptará a través de sus coeficientes. Esta matriz, por tanto, hace que cualquier formato previo o cualquier mezcla de órdenes en nuestro campo sonoro se pueda adaptar a la perifería esperada en Ambisonics. Teóricamente es posible incluso acomodar campos sonoros que únicamente abarquen subconjuntos de la esfera.

Capítulo 4

Aplicación del Ambisonics en Grabación y Reproducción

Una vez vista y repasada la tecnología Ambisonics, es hora de llevarla a la práctica y montar diferentes configuraciones para su grabación y reproducción. Al contrario de lo que cabría esperar, en primer lugar se realizó el diseño de los sistemas de reproducción, para los cuáles se utilizó material Ambisonics alojado en la red para su testeo. De esta manera, se pudo realizar el trabajo previamente sin tener que poseer material Ambisonics propio.

Para llevar a cabo la parte práctica del presente proyecto se han utilizado materiales e instalaciones facilitadas por el GTAC (Grupo de Tratamiento de Audio y Comunicaciones) de la UPV.

4.1 Reproducción

Una vez vista la tecnología Ambisonics, y antes de repasar algunas técnicas para la grabación de material Ambisonics, nos centraremos en desarrollar y buscar diferentes métodos de reproducción, para los cuáles se utilizará material Ambisonics disponible en la red.

En primer lugar, y antes de poder trabajar en el laboratorio de pruebas, se llevó a cabo la realización de una serie de desarrollos para la escucha de material *surround* y Ambisonics mediante escucha binaural, es decir, con auriculares. Para ello, se llevó a cabo la codificación de un sistema de escucha 5.1 en formato binaural a través de Ambisonics, de manera que cualquier material con audio 5.1 podrá ser escuchado en 3D por escucha binaural gracias a Ambisonics.

En segundo lugar, se llevó a cabo la realización de un proyecto en Arduino para la construcción de un *Head-tracker* de bajo coste. De esta forma, podremos utilizar nuestro seguidor para la escucha en 3D de material Ambisonics, pudiendo detectar el giro de la cabeza y aplicar el movimiento de la escena sonora en consecuencia.

En tercer lugar se implementaron una serie de decodificadores Ambisonics y sus arreglos físicos de altavoces en el laboratorio de "Percepción Espacial del Sonido" para la reproducción de material Ambisonics de distinto orden.

4.1.1 5.1 binaural

En la presente sección se va a explicar, paso a paso, como virtualizar un sistema de escucha envolvente 5.1 a través de Ambisonics para su decodificación a escucha binaural. De esta forma, el usuario podrá disfrutar, sin necesitar de un sistema “home cinema” físico, de una reproducción envolvente 5.1 a través de auriculares.

El software empleado para dicha virtualización es:

- **Rogue Amoeba Loopback 2:** software de pago para la generación de interfaces virtuales de audio. Nos permitirá crear el ruteo interno de la señal entre el software de reproducción de vídeo y el software de procesamiento de audio Ambisonics.
- **Reaper:** se utilizará como host VST para cargar la cadena de procesado de los 6 canales de audio provenientes del vídeo.
- **SPARTA AmbiENC y AmbiBIN:** AmbiENC se utilizará para la codificación de los 6 canales del vídeo a formato Ambisonics de tercer orden, mientras que AmbiBIN se encargará de la decodificación del formato Ambisonics a binaural.
- **VLC:** Reproductor de vídeo gratuito y *open source* con multitud de códecs tanto de vídeo como de audio.

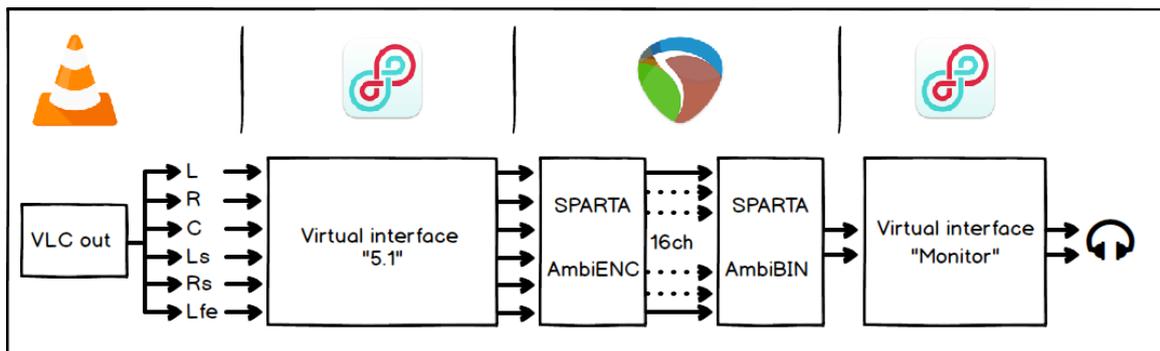


Figura 4.1: Ruteo de señal para virtualizar un sistema 5.1.

Como se observa en la figura 4.1, se trata de una cadena separada en 4 partes bien diferenciadas. A continuación se expone con detalle la configuración interna de cada una de dichas partes para su correcto funcionamiento. Si bien se observa que la primera parte de la cadena empieza a través de VLC, en primer lugar habrá que crear las interfaces de audio virtuales en Loopback 2.

Loopback 2

En primer lugar deberemos crear dos interfaces virtuales a través del software Loopback 2, de Rogue Amoeba. Se trata de un software de pago, siendo elegido por falta de una alternativa gratuita conocida¹.

Se deberán crear dos interfaces virtuales, una encargada de encaminar los 6 canales de audio originales del vídeo con audio 5.1 hacia Reaper, software donde se llevará a cabo el procesado. La segunda, será la encargada de sacar la salida de audio desde Reaper hacia nuestra salida de auriculares, desde donde poder monitorizar los resultados.

¹A fecha 11 de Diciembre de 2019 se descubrió la existencia de una versión nueva y gratuita denominada *BlackHole* y disponible en GitHub. Se continuó con el uso de Loopback 2 por tenerlo ya comprado y configurado correctamente.

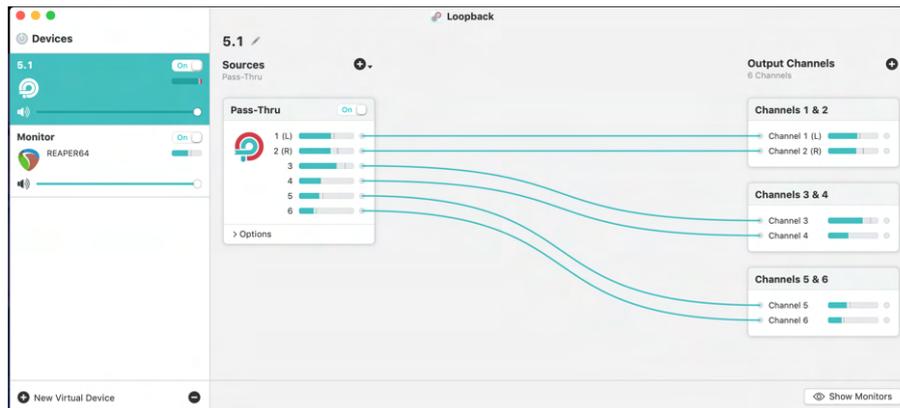


Figura 4.2: Interface virtual de audio 5.1.

Tal y como se observa en la figura 4.2, se ha creado un nuevo dispositivo llamado “5.1”, para el cuál se han creado 6 salidas, y utilizando la opción *Pass-Thru* de Loopback (con esta función, el audio que internamente está siendo reproducido por el ordenador, aparece como una fuente de audio), se dirigen los 6 canales de audio a sendas salidas virtuales.

Para llevar a cabo este punto, se configuró el orden de los canales desde la aplicación “Configuración de Audio MIDI”, de macOS, tal y como se observa en la figura 4.3.



Figura 4.3: Configuración del orden de canales del sistema 5.1.

Como dentro de Reaper deberemos seleccionar la interface de audio virtual “5.1” para la captura de audio, deberemos crear una segunda interface virtual para monitorizar la salida de Reaper por nuestros auriculares. Para ello se ha creado una segunda interface virtual denominada “Monitor”, la cuál toma los dos canales de salida de Reaper y los conduce a la salida física de nuestro ordenador, donde conectaremos nuestros auriculares para disfrutar de una escucha binaural.

Tal y como observamos en la figura 4.4, la fuente “Pass-Thru” se desactiva (si se dejase activada, por ella sonaría además de la reproducción de Reaper, los dos primeros canales de VLC, ya que estos estarían siendo reproducidos simultáneamente). De esta manera estaremos ruteando la salida binaural de Reaper a la salida física de nuestro ordenador.

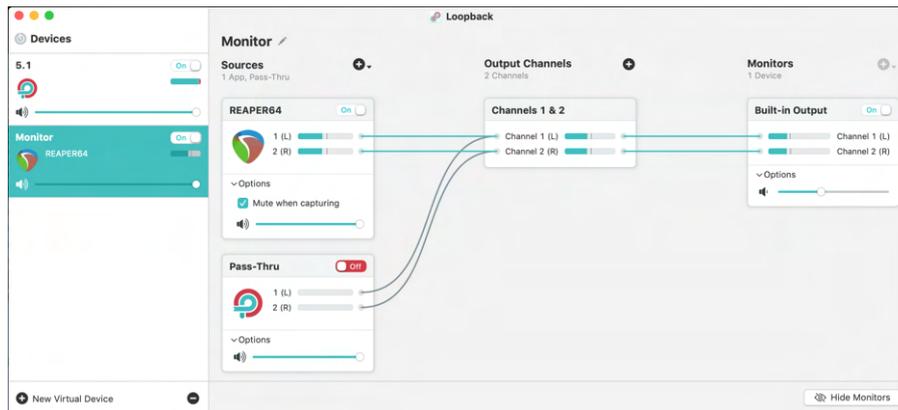


Figura 4.4: Interface virtual de audio para monitorización de la salida de Reaper.

VLC

Dentro de VLC la configuración necesaria es básica. En primer lugar, y ya con las interfaces virtuales creadas, deberemos seleccionar la interface “5.1” como interface de salida.

En segundo lugar, y si no estuviera configurado, deberemos entrar a las preferencias, y tal y como se observa en la figura 4.5, activar la pestaña “Forzar soporte S/PDIF”, así como el modo de salida de audio en modo Dolby Surround.

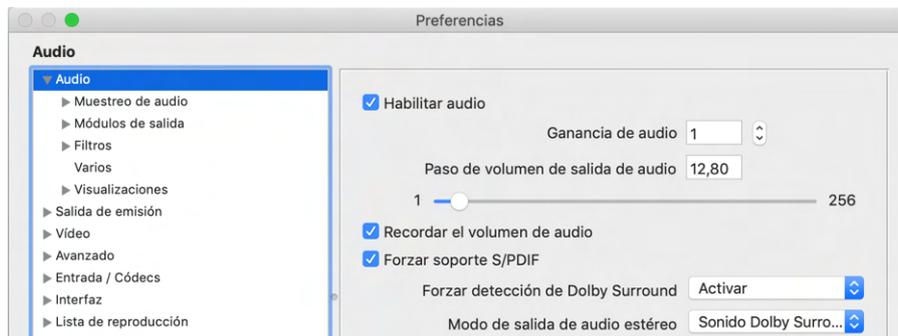


Figura 4.5: Configuración interna de VLC para la reproducción en 5.1.

Una vez configurado VLC con la interface de salida de audio seleccionada, únicamente nos faltará crear la sesión interna de Reaper para el procesamiento del audio.

Reaper

Reaper es realmente el núcleo de todo este proceso y donde realmente se lleva a cabo el procesamiento de audio. La decisión de utilizar Reaper como host para cargar los plugins de procesado Ambisonics se debe a su carácter “gratis” y flexible, donde prácticamente cualquier ruteo de señal es posible, siendo mucho más flexible que otros DAW.

Una vez en Reaper, lo primero que debemos hacer es ir a la pestaña REAPER → Preferencias → Audio → Dispositivos de audio y ahí seleccionar como interface la que hemos creado en Loopback llamada “5.1”. Con la interface de audio seleccionada, deberemos crear una pista con la configuración de ruteo que se observa en la figura 4.6. Es importante activar el check de monitorización de audio, representado en la pista como una letra “I” en color verde. Además, en la pestaña de ruteo de pista,

deberemos desactivar el envío máster y deberemos elegir el número de canales de la pista, en este caso 6, y realizar un envío de los 6 canales a una segunda pista que llamaremos “BUS AMBI”.



Figura 4.6: Bus de Reaper para la captura del audio procedente de VLC.

Ya con el segundo bus creado y con las pistas ruteadas hacia éste, es aquí donde se hará el procesado mediante una cadena de plugins VST. La configuración de ruteo de esta pista será la que se observa en la figura 4.7, con el envío a máster desactivado, y una salida física hacia las salidas 1 y 2, correspondientes a la salida del propio Reaper hacia la segunda interface virtual creada en Loopback. Además, es importante marcar que el número de canales para esta pista será de 16, ya que trabajaremos en Ambisonics de tercer orden.



Figura 4.7: Bus de Reaper para el procesado de audio en Ambisonics.

Una vez con la pista creada y configurada, realizaremos una cadena de procesamiento VST compuesta por AmbiENC → AmbiBIN.

- **AmbiENC:** En este punto realizaremos la codificación de 5.1 a Ambisonics. Para ello hemos creado, según la normativa **UIT-R BS.775-1**[13] la configuración 5.1 según el orden de los altavoces configurados en 4.3. De esta manera, cada uno de los canales corresponderá con el altavoz físico que representa en 5.1, y a su vez, ocupará su posición física en el campo de sonido Ambisonics que hemos codificado. Además, el orden elegido para la codificación ha sido tercer orden, el cuál necesita de 16 canales para representar todo el campo sonoro. Se ha elegido tercer orden por ser una situación de compromiso entre una buena resolución espacial y una carga de procesamiento adecuada para mi ordenador y en definitiva para la mayoría.
- **AmbiBIN:** Una vez hemos convertido los 6 canales surround a Ambisonics con AmbiENC, es el momento de decodificar el espacio sonoro creado a formato binaural para una correcta monitorización mediante auriculares. Para ello vamos a utilizar el decodificador AmbiBIN de la suite SPARTA. En él, podemos elegir desde el orden Ambisonics, tipo de codificación de los canales Ambisonics, e incluso podemos cargar nuestra HRIR para corrección binaural en formato SOFA. Tras este punto, los dos canales en formato binaural que salen de AmbiBIN corresponderán al 5.1 inicial codificado en Ambisonics binaural, y por tanto se rutearán hacia

fuera de Reaper mediante Loopback para su monitorización, tal y como ya se ha explicado previamente.



Figura 4.8: AmbiENC, de la suite Sparta Ambisonics. Codificador Ambisonics.

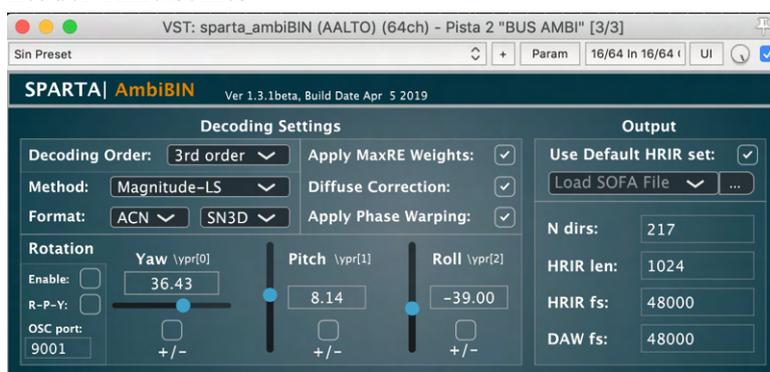


Figura 4.9: AmbiBIN, de la suite Sparta Ambisonics. Decodificador binaural Ambisonics.

4.1.2 *Head-tracker* DIY

Teniendo un campo sonoro tridimensional como el que nos ofrece Ambisonics, la posibilidad de poder girar la cabeza y que el campo sonoro gire como si estuviéramos en él resulta atractiva. A este tipo de dispositivo se le conoce como *head-tracker*, y se encarga de realizar un seguimiento de los movimientos de la cabeza mediante un acelerómetro y un giroscopio. Para ello existen multitud de soluciones comerciales como el de la conocida marca de hardware y software para audio *Waves*, además de otra serie de soluciones como las que ofrecen las casas *Track-hat* o *DelanClip*.

Sin embargo, a continuación se va a proceder a explicar como llevar a cabo el desarrollo y la implementación de un *head-tracker* propio basado en Arduino a partir del desarrollo de *Tomasz Rudzki* y *Jacek Majer* en GitHub [14]. Para ello, únicamente se necesita una placa Arduino Pro Micro, la placa sensora MPU-9250 y el cableado para las placas además de un cable USB para la conexión con el PC.

- **Arduino Pro Micro.**- Placa equipada con el chip ATMEGA32U4 con interfaz USB incorporada. Dispone de una alimentación de 5V a través de la propia conexión USB, además de un

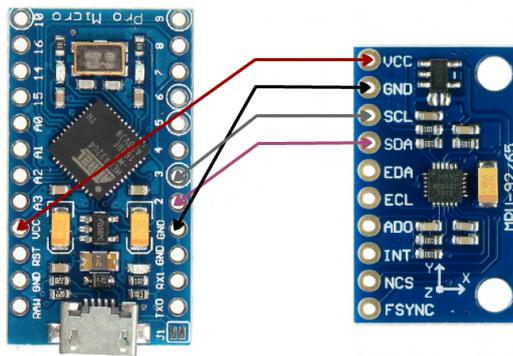


Figura 4.10: Esquema de conexionado entre la placa Arduino Pro Micro y la placa sensora.

reloj interno de 16MHz. Además, es posible extraer 5V por su pin Vcc para alimentar el MPU-9250, lo que lo convierte en un todo-en-uno perfecto para este propósito por no depender de complementos externos más allá de la placa sensora. Su precio en internet ronda los 8€.

- **MPU-9250.-** Se trata de una IMU² de 9 DoF³ equipada con acelerómetro, giroscopio y magnetómetro. Es por tanto una IMU de nueva generación capaz de medir aceleración lineal, velocidad de rotación angular, y vectores de campo magnético. Su precio en internet ronda los 10€.

Siguiendo el conexionado que se muestra en la figura 4.10, flasheamos el software disponible en GitHub en nuestra placa. Además, deberemos descargar el software que hace de puente entre la placa y nuestro DAW, también conocido como *bridge*, también disponible en GitHub. Una vez realizadas las conexiones de alimentación (VCC), masa (GND), datos (SDA) y reloj (SCL), y cargado el programa en la placa, deberemos seguir los pasos para la configuración adecuada de nuestro DAW con el *head-tracker*.

En el presente documento no se van a entrar en más detalles sobre la construcción del *head-tracker* debido a que estos pasos ya vienen suficientemente detallados en el post original de los autores. Sin embargo, donde sí que vamos a entrar en detalles es en la configuración y uso de nuestro *head-tracker* para trabajar con Ambisonics.

En concreto, vamos a explicar como podemos sincronizar nuestro proyecto Ambisonics dentro de nuestra sesión con el *head-tracker*. Para ello, haremos uso de la sesión creada para la reproducción de un sistema 5.1 virtual mediante Ambisonics explicada en la sección 4.1.1, en donde añadiremos la funcionalidad de poder colocarnos nuestro *head-tracker* en los auriculares y con nuestros movimientos de cabeza, sentiremos como si nos moviésemos acústicamente dentro de la escena sonora.

Para ello, y con el *head-tracker* funcionando dentro de nuestro DAW mediante el protocolo OSC, deberemos añadir un nuevo plugin a la cadena, el SceneRotator de la suite IEM. La cadena quedará *AmbiEnc*→*SceneRotator*→*AmbiBin*, de forma que codificaremos el sistema 5.1 a Ambisonics, en este dominio realizaremos las rotaciones necesarias en tiempo real para girar la escena a través del *head-tracker* y el resultado se decodificará a binaural para su escucha por auriculares.

- **SceneRotator.-** Gracias a SceneRotator es posible girar en sus tres ejes el campo sonoro Ambisonics de manera manual. Además, este plugin acepta comunicaciones tanto MIDI como OSC,

²IMU.- Unidad de medición inercial multipropósito.

³9 DoF indica los grados de libertad de la unidad de medición.

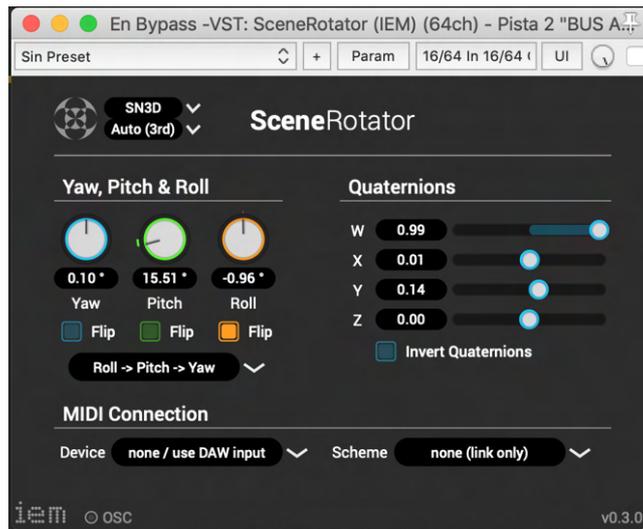


Figura 4.11: SceneRotator de la suite IEM para la rotación del campo sonoro Ambisonics.

dando la posibilidad al usuario de controlarlo externamente. En nuestro caso deberemos, mediante el *bridge* de nuestro *head-tracker*, indicarle cada una de las componentes de la placa a que tipo de giro corresponde (Yaw, Pitch y Roll). Para una correcta configuración se recomienda seguir el tutorial de *Gabriel Zalles*.^[15]

Una vez tengamos configurados tanto Reaper, como el plugin SceneRotator para funcionar con nuestro HeadTracker, únicamente deberemos acoplarlo a la diadema de nuestros auriculares y mediante los giros de nuestra cabeza veremos como SceneRotator traduce esos movimientos físicos en giros del campo sonoro.

Si bien se ha explicado como utilizar el *head-tracker* para visualizar películas 5.1 en binaural, el *head-tracker* es útil para cualquier mezcla en formato Ambisonics que se nos ocurra. Para ello únicamente deberemos colocar los plugins SceneRotator y AmbiBin en nuestro bus Ambisonics de salida para una correcta monitorización 3D con seguimiento de la cabeza.

4.1.3 Laboratorio de pruebas

El laboratorio donde finalmente se llevaron a cabo las pruebas e implementación de nuestro **sistema de reproducción Ambisonics** fue el laboratorio de “Percepción Espacial del Sonido” perteneciente al departamento GTAC del iTeam, situado en el bloque N de la Ciudad Politécnica de la Innovación (CPI) de la UPV.

Se trata de un laboratorio concebido y destinado tanto a la reproducción de material WFS⁴ como a la medición de la HRTF personal. Para tal fin, el laboratorio se compone de un sistema de reproducción 360° compuesto por tres alturas de reproducción repartidas en tres anillos. El anillo central es el principal y ofrece una resolución de 5°, mientras que los anillos superior e inferior ofrecen una resolución de 45°. Inicialmente, el propósito de este laboratorio no era el de servir como sistema modular para reproducir material Ambisonics, pero debido a su versatilidad gracias al gran número de altavoces del que se compone, concretamente 88, se ha elegido por las posibilidades que ofrece a la hora de poder implementar distintos arreglos en un sólo laboratorio.

⁴WFS.- *Wave Field Synthesis* es una técnica de renderización de audio espacial caracterizada por la creación de entornos y fuentes acústicas virtuales. ^[33]

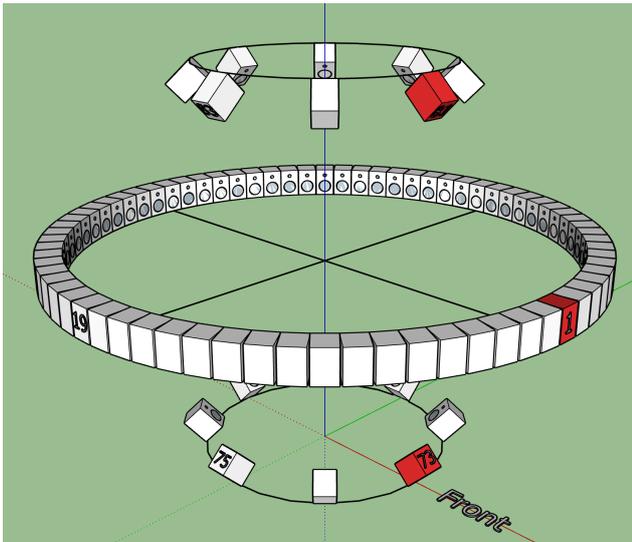


Figura 4.12: Modelo 3D del laboratorio de pruebas levantado en *SketchUp*.



Figura 4.13: Laboratorio de "Percepción Espacial del Sonido".

Se puede observar en la figura 4.12 un modelo a escala levantado en el software *SketchUp*, donde los altavoces marcados en rojo indica el altavoz correspondiente a la posición de 0° . Además, los números marcados en los altavoces indican el número de altavoz dentro del sistema. Esta numeración se comentará detalladamente mas adelante.

El conjunto está formado, tal y como se ha dicho, por 88 altavoces repartidos en 3 anillos. El anillo central, el más grande y con 72 altavoces, uno cada 5° , se encuentra a una altura desde la base de cada altavoz al suelo de 1,60 metros, y tiene un diámetro de 4 metros desde el frontal de cada altavoz al frontal del altavoz opuesto. El anillo inferior y superior tienen un diámetro de 2 metros y poseen 8 altavoces cada anillo, uno cada 45° . El anillo inferior se encuentra a ras de suelo y cada altavoz tiene una inclinación hacia el centro de la esfera de 45° desde la base de cada altavoz al suelo. El anillo superior por su parte tiene la misma inclinación desde el techo a su parte superior, 45° , y se encuentra a 1,6m desde el techo a la base de los altavoces del anillo central. En el Anexo B se pueden observar las coordenadas esféricas de cada altavoz dentro del sistema, donde el centro viene representado por el centro del anillo central, donde se encuentra el *sweet-spot*. Es importante tener presente estas coordenadas ya que nos serán necesarias a la hora de preparar los distintos decodificadores AllRAD para Ambisonics.

Material

El laboratorio, tal y como se ha dicho, está compuesto de 88 altavoces repartidos en 3 anillos, para los cuáles son necesarias 4 tarjetas de sonido externas y una tarjeta PCI instalada en el ordenador con el que trabajemos para el control de cada una de las 4 tarjetas. En la figura 4.19 se puede observar un diagrama de conexionado de todo el laboratorio. En la siguiente sección se verán las partes importantes del sistema, dejando a un lado materiales secundarios como el cableado, soportes de montaje, etc.

- **M-Audio BX5 D2**

Los altavoces, todos del modelo M-Audio BX5 D2, son bi-amplificados, es decir, en su interior albergan tanto los dos transductores (woofer de kevlar de 5" y tweeter de cúpula de seda de 1"), como un amplificador dedicado a cada una de las dos vías y el filtro correspondiente. Esto hace que no sea necesario alimentar los altavoces con 88 etapas de potencia externas, aunque ofrece el inconveniente de que a cada altavoz hay que llevar tanto un cable de señal como uno de alimentación, si bien al ser una instalación fija esto no resulta un problema. A continuación se ofrece una pequeña tabla de especificaciones:

Especificaciones	
Respuesta en frecuencia:	56 Hz - 22 kHz
Frecuencia de corte:	3 kHz
Potencia amplificador de graves:	40 Watts
Potencia amplificador de agudos:	30 Watts
SNR (Rango dinámico):	>100 dB
Impedancia de entrada:	20 kOhms (XLR); 10 kOhms (TRS)
Sensibilidad de entrada:	85 mV de ruido rosa producen 90 dBA a 1 metro
Tamaño:	19.5 cm; 17.6 cm; 25.2 cm
Peso	5 kg

Como observamos en la tabla de especificaciones y en la figura 4.14, el altavoz posee dos posibilidades a la hora de hacerle llegar la señal, por XLR balanceada, o por TRS/Jack de 6,3mm no balanceada. En nuestro caso, el laboratorio está totalmente instalado con cableado XLR balanceado para reducir así el posible ruido ocasionado por las tiradas largas de cableado.

Es importante indicar que todos los altavoces del laboratorio están configurados con su volumen en el punto intermedio, es decir, a un 50 % de la ganancia de su etapa de amplificación. Debido al gran número de altavoces, con esa ganancia será suficiente, ya que el volumen general se controlará vía software desde nuestro DAW y nos evitaremos llevar al amplificador de cada altavoz a una zona de trabajo no lineal, donde pueda añadir distorsiones no deseadas.



Figura 4.14: Altavoces *M-Audio-BX5 D2*.

<https://m-audio.com/products/view/bx5-d2>

- **Motu 24I/O**

Para alimentar 88 altavoces con señales independientes entre sí, es necesario por tanto disponer de 88 salidas analógicas desde nuestra tarjeta de sonido para poder alimentar cada uno de los altavoces. 88 salidas no es algo sencillo de abordar, debido en parte a que la mayoría de los equipos comerciales no ofrecen una cantidad tan elevada. Para dar solución a dicho problema, en el laboratorio hay instalados cuatro módulos *Motu 24I/O*, cada uno de ellos con 24 entradas y 24 salidas analógicas tipo TRS, así como sus correspondientes conversores AD/DA. Cada uno de estos módulos posee además una interfaz de comunicación digital AudioWire, propietaria de MOTU, y que utiliza la interfaz física Firewire IEEE 1394 para su comunicación con el PC a través del sistema Motu PCI-424, el cual se explicará en detalle a continuación y que ofrece la posibilidad de conectar hasta 4 módulos Motu compatibles para que operen virtualmente como una sola tarjeta, ofreciendo así un gran número de entradas y salidas con su configuración de 4 módulos, concretamente 96 entradas y 96 salidas. Esto hace que el 24I/O no sea una tarjeta autónoma en sí misma, si no un módulo del sistema Legacy PCI de Motu, ya fuera de catálogo.

Se trata de una tarjeta algo obsoleta, por lo que la información disponible de la misma es algo limitada, así como el hecho de trabajar en AudioWire y necesitar de la PCI-424 para trabajar, hace que no sea una solución versátil ni práctica para la mayoría de los sistemas. Por otra parte, es importante mencionar que difícilmente necesitaremos un sistema con 88 salidas si no es para una instalación similar al laboratorio de pruebas.



Figura 4.15: MOTU 24I/O.

Cada una de los módulos ofrece por tanto 24 entradas y 24 salidas analógicas de tipo TRS, la posibilidad de trabajar en frecuencias de muestreo seleccionables por el usuario de 44.1, 48 o 96 kHz, comunicación AudioWire con interfaz física Firewire, wordclock y vúmetros digitales de control de 5 segmentos para las 24 entradas y las 24 salidas de la tarjeta. A continuación se expone una pequeña tabla con sus especificaciones principales:

Especificaciones	
Número de entradas/salidas:	24/24 +4dB TRS balanceadas
Frecuencia de muestreo:	44.1, 48 o 96 kHz
Profundidad de bits:	24 Bits
<i>Oversampling:</i>	x128
Extras:	Wordclock, vúmetros en todas las entradas/salidas.
Interface:	AudioWire a MOTU PCI-424

- **Motu PCI-424/PCIe-424**

Debido a la naturaleza del sistema implementado, y tal y como se ha comentado, la tarjetas MOTU 24I/O no son tarjetas autónomas por sí mismas, ya que la verdadera interfaz de comunicación con el PC es la tarjeta PCI-424/PCIe-424, de Motu, la cuál actúa como cerebro DSP de un sistema modular de hasta 4 elementos, los cuáles actuarán virtualmente como un único sistema mayor. Para ello, Motu ofrecía los módulos 2408mk3, HD192 y 24I/O. Cada uno de ellos ofrecía

una combinación diferente de entradas salidas, posibilidad de incluir previos de micrófono, etc, de manera que el usuario podía elegir qué módulos necesitaba para su configuración, y mediante la PCI, juntarlos en un sistema único.

En el caso del laboratorio, y tal y como se ha visto, la combinación elegida busca maximizar el número de salidas, de manera que 4 Motu 24I/O fueron las elegidas. Además de los 4 puertos AudioWire, la tarjeta cuenta con un puerto multi pin para sincronía ADAT con otros dispositivos externos, así como un procesador DSP con el software CueMix DSP para mezcla y monitoraje de las entradas y salidas del sistema. Por último incorpora compatibilidad con drivers WDM y ASIO para Windows.



Figura 4.16: MOTU PCI-424.

• PC

Para el montaje de nuestro sistema, se pensó inicialmente en el uso del ordenador dispuesto en el rack del laboratorio, y que usualmente se utilizaba para sistemas WFS y medición de la HRTF. Sin embargo, por motivos que no se han llegado a constatar, éste no era capaz de reconocer los VST instalados, aún siendo formateado y repasado a fondo, por lo que finalmente se descartó su uso a pesar de su potencia. Como solución, se instaló un PC externo al laboratorio y se le instaló la tarjeta PCI-424 para la comunicación con el resto de sistemas.

El software instalado para trabajar específicamente con Ambisonics fue el DAW Reaper y los paquetes VST de SPARTA/COMPASS de la Aalto University, así como el paquete IEM del Institute of Electronic Music and Acoustics. La idea es que con estos dos paquetes VST de tratamiento de audio Ambisonics, así como el software propio de cada elemento del sistema que sea necesario utilizar, como convertidores A/B para los micrófonos, software CueMix DSP de las tarjetas, drivers, etc, será todo el software utilizado. De esta manera, se va a utilizar únicamente software gratuito y al alcance de todos, así como los distintos software ligados a los elementos hardware que se vayan a utilizar.

Por tanto, y una vez explicado, a continuación se exponen las características principales del PC utilizado en el laboratorio:

Especificaciones	
Procesador:	Intel Core i7-3770 @ 3,40GHz
Núcleos/Hilos:	4/8
Memoria RAM:	8 GB
SO:	Windows 7 Professional 64 Bits
HDD1:	298 GB
HDD2:	372 GB

Flujo de señal

Una vez vistos los materiales, se va a proceder a explicar el ruteo, o flujo de señal a nivel hardware, desde que sale del PC a través de la salida AudioWire correspondiente, hasta que suena por el altavoz.

Para ello podemos observar la figura 4.19. En primer lugar observamos como en el PC tenemos la tarjeta PCI-424 con sus 4 puertos AudioWire numerados de izquierda a derecha. El número correspondiente va a asociado al orden interno en el que se reconocerán las tarjetas, por lo que se deben conectar en dicho orden, tal y como se observa en cada una de las tarjetas, donde también están numeradas. Para ello, deberemos hacer uso de 4 cables Firewire 400 para la comunicación digital entre el ordenador y los módulos 24I/O.



Figura 4.17: Vista frontal del rack con las 4 tarjetas MOTU 24I/O.



Figura 4.18: Vista trasera del rack con las 4 tarjetas MOTU 24I/O y el cableado necesario.

Tras este paso, la tarjeta realiza las conversiones D/A correspondientes y saca en formato analógico la señal a cada uno de los altavoces. Para un control adecuado de cada uno de los altavoces, éstos están numerados de la siguiente forma:

Numeración de los altavoces:	
Anillo intermedio:	1(0°) - 72(5°)
Anillo inferior:	73(0°) - 80(45°)
Anillo superior:	81(0°) - 88(45°)

Sin embargo, a pesar de que la numeración de los altavoces se realiza en sentido horario porque así estaba determinada en el montaje original, el estándar SOFA nos indica que la graduación de los grados para las determinadas posiciones va en sentido anti horario, tal y como viene indicado en [34], capítulo 3.7.2. Por tanto, y teniendo un anillo central de 72 altavoces y dos de 8, se dedicarán las 3 primeras tarjetas MOTU a proveer las salidas analógicas para alimentar el anillo central, dejando la última tarjeta para alimentar los dos anillos en altura. Existe una excepción, tal y como se observa en la imagen con el flujo de señal, donde la salida 1 de la segunda tarjeta, es decir, la correspondiente al altavoz 25, no funcionaba, por lo que se tuvo que re colocar en el primer canal libre de la cuarta tarjeta, es decir, el 17 de la misma. Todo esto viene indicado en el diagrama, donde los números en

color rojo de los anillos indican el número de altavoz, mientras que los números de color provenientes de cada tarjeta, indican el canal correspondiente a dicha tarjeta.

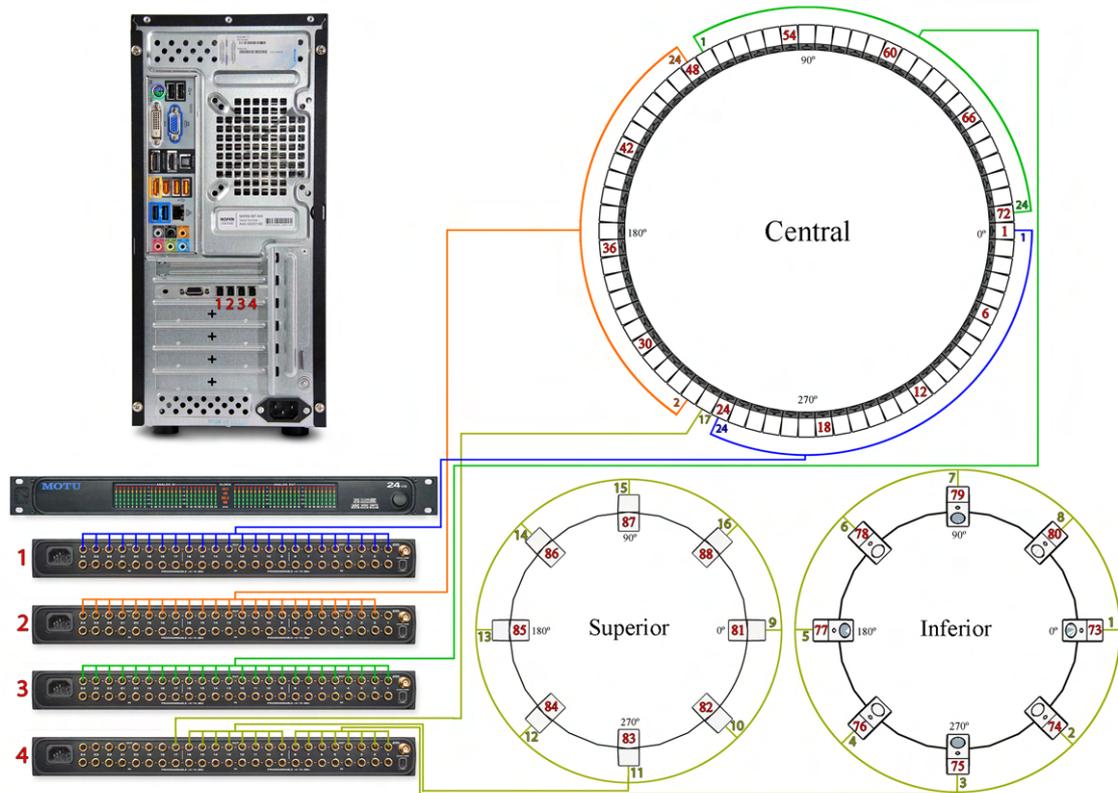


Figura 4.19: Diagrama de conexionado del laboratorio.

A pesar de estar todo numerado, se hace complicado a la hora de querer montar un arreglo concreto, en parte por la gran cantidad de altavoces, y en parte por la numeración de éstos. Al estar conectados en sentido horario, pero estar graduados en sentido anti horario, y además con sus conexiones divididas en 4 tarjetas separadas, cuando queremos hacer el ruteo interno desde nuestro DAW, esta se vuelve una tarea complicada. Para ello se programó una sencilla hoja XLS con macros (importante habilitarlos) para la tarea de de-codificación.

Decodificador de arreglos Ambisonics

Para llevar a cabo la tarea de decodificar cualquier arreglo Ambisonics a nuestro laboratorio se ha desarrollado una pequeña hoja .xls(Excel) para dicha tarea. Dicha hoja se ha programado con macros en VBA(Visual Basic for Applications). En dicha hoja se han representado 4 bloques que representan cada una de las 4 tarjetas MOTU (1-Azul, 2-Naranja, 3-Verde y 4-Amarillo). En cada bloque podemos observar el número de tarjeta, sus 24 canales, el anillo físico al que alimentan, así como el número de altavoz de cada canal de la tarjeta y el ángulo que representa. Además, en color rojo y negrita se muestra la excepción del altavoz 25, cuya salida no es la que correspondería de forma lógica porque ésta falla, tal y como ya se ha comentado, por lo que viene representado de forma clara.

En la parte inferior se observan dos bloques bien diferenciados, uno para decodificar en formatos Surround (5.0, 7.0, 9.0 y 11.0) y otro para formatos Ambisonics (1^{er} - 7^o orden).

Para el caso de los formatos Surround, se han seguido los estándar descritos por Dolby en [35], si

bien únicamente se han decodificado las configuraciones para el plano horizontal y superior debido a que Dolby no contempla altavoces en la parte inferior de la escucha. Además, tampoco se ha incorporado el canal LFE destinado a bajas frecuencias, ya que al tratarse de un sistema para pruebas, y no para reproducción, no se ha considerado necesario. Su incorporación sería sencilla al tratarse de un único canal.

Para el caso de los formatos Ambisonics, estos se han obtenido en combinación con lo descrito en [36] y a partir de la imagen 3.1, a partir de la cuál se puede deducir que para cada orden, existe una separación entre los lóbulos Ambisonic. Siguiendo esto, es lógico pensar que la separación de cada uno de esos lóbulos se podrá colocar un altavoz, de forma que tengamos un altavoz dedicado a cada uno de los lóbulos. De esta forma nos aseguramos el número justo de altavoces para cada orden, evitando el exceso de altavoces que nos puedan provocar problemas de filtrado o reducción del conocido *sweet-spot*. A continuación se expone una tabla con los distintos arreglos Ambisonic implementados, únicamente en el anillo horizontal:

Orden	Nº altavoces	Separación (°)
1	4	90
2	6	60
3	8	45
4	9	40
5	12	30
6	18	20
7	24	15

Por otro lado, se ha implementado la posibilidad de rotar o añadir un *offset* a estos arreglos. Esto quiere decir que todos parten del altavoz frontal, el número 1, y se reparten equidistantes a partir de ese origen. La posibilidad de rotar hace que el origen ya no sea el número 1, de manera que es como si girásemos los lóbulos Ambisonics. No es una opción muy utilizada, pero se ha creído conveniente añadirla para ofrecer dicha alternativa.

Figura 4.20: Excel para decodificar arreglos al laboratorio.

Por último, tenemos la posibilidad de añadir altavoces tanto en el anillo superior como en el inferior. Estos arreglos vienen por defecto limitados a 4 u 8 altavoces por anillo, ya que éstas son las dos únicas posibilidades que se pueden adaptar al laboratorio. Se han elegido 4 altavoces para los órdenes que van del 1 al 4, y del 5 al 7 son 8 altavoces por altura.

Por tanto, el usuario elige el tipo de decodificador deseado, pudiendo añadir altura u *offset* si lo desea, y a continuación calcular el decodificador. Cada altavoz que forme parte del arreglo se iluminará en color rojo (anillo central), azul (inferior) o verde (superior), indicando al usuario qué altavoces tendrá que configurar en el ruteo interno del DAW como salida. De esta manera, el usuario sabrá en todo momento qué altavoces necesita, los ángulos que representan, la tarjeta MOTU que los mueve, etc.

Tal y como se observa en la figura 4.20, en ella se ha preparado un arreglo de tercer orden con altavoces tanto en altura superior como inferior. El programa entonces indica al usuario que los altavoces que se deben utilizar son el 1, 10, 19, 28, 37, 46, 55 y 64 para el anillo central, y los correspondientes para el superior e inferior.

De esta forma, se hace mucho más sencillo para quien desee preparar los distintos tipos de arreglos en el Laboratorio de pruebas. Por supuesto, la posibilidad de añadir cualquier otro decodificador es bastante sencilla a través de VBA.

Se pueden observar todos los sistemas implementados en el Anexo C.

4.2 Grabación

Inicialmente, se pensó para el siguiente proyecto una serie de grabaciones a partir de los micrófonos Ambisonics disponibles en el departamento GTAC de la UPV. Concretamente, los micrófonos *Rode NT-SF1* y *TSL Soundfield ST450 MKII*, ambos de primer orden, y el *Eigenmike em32* de hasta cuarto orden. Por problemas derivados de la pandemia de la COVID-19, y con el cierre de los laboratorios, dichas grabaciones se volvieron imposibles de llevar a cabo. Es por ello que en esta sección se va a proceder a explicar, de manera teórica y a modo de manual, cómo llevar a cabo grabaciones con cada uno de estos 3 micrófonos y grabaciones Ambisonics en general, posibilidades que ofrece cada uno, así como sus ventajas y desventajas en cada situación.

Para ello se va a proceder a explicar el montaje y la utilización general para grabaciones Ambisonics, las cuáles serán aplicables a cualquier micrófonos. A medida que vayamos avanzando por cada uno de los puntos clave, se profundizará en los detalles particulares y características propias a tener en cuenta para cada uno de los tres micrófonos mencionados. De esta forma, se verá tanto el flujo de trabajo para una grabación Ambisonics estándar, como la focalización de la misma para los micrófonos disponibles.

Para explicar este punto, se va a realizar un repaso al flujo de trabajo habitual, el cuál se puede observar en la figura 4.21, empezando por la situación y colocación del propio micrófono, pasando a continuación a la interfaz asociada al mismo (si es que la hubiera) y por último viendo posibles soportes de grabación adecuados para grabación en Ambisonics, tanto de escritorio como portátiles en general, y para cada uno de estos micrófonos en particular.

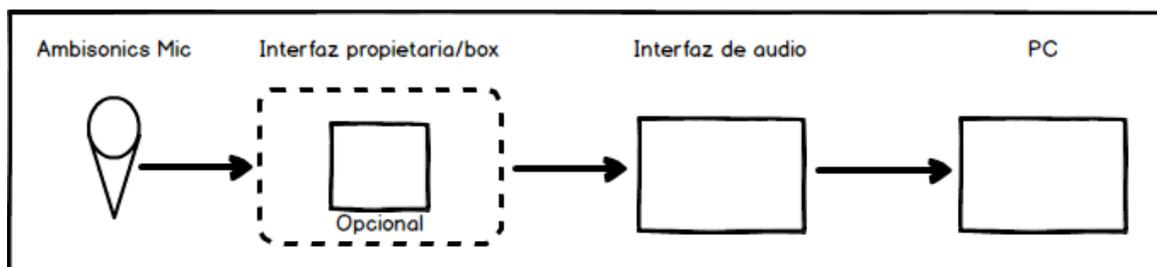


Figura 4.21: Flujo habitual de grabaciones Ambisonics.

4.2.1 Situación del micrófono

Uno de los primeros puntos que debemos tener en cuenta a la hora de llevar a cabo una grabación Ambisonics es la situación del micrófono. Cómo y dónde se va a colocar es de suma importancia, ya que de ello dependerán factores tan importantes como la situación de las fuentes de importancia en el campo sonoro, nivel y ubicación de las mismas en la escucha final, etc. Si bien es cierto que uno de los potenciales de las grabaciones Ambisonics es la posibilidad de alterar en post producción el campo sonoro, pudiendo focalizar la escucha en un punto dentro del espacio, eliminar ruido proveniente de direcciones no deseadas, etc, también es cierto que una mala captación difícilmente se podrá arreglar en post producción. Es por todo lo comentado que la situación del micrófono cobra una relevancia especial y debemos ser muy cuidadosos con la misma.

Suspensión y/o anti viento

Uno de los puntos iniciales para la grabación es la suspensión del micrófono. Debemos ser cuidadosos con este punto ya que una suspensión rígida, tipo pinza, como las que suelen venir incluidas en los micrófonos de fábrica, a menudo nos va a dar resultados muy pobres en temas de vibraciones. Cualquier golpe, pisada o similar que reciba el suelo, así como vibraciones del cableado, se convertirán en ruido en nuestra grabación. Es por ello que se hace necesario el uso de suspensiones tipo araña o similar, capaces de aislar el micrófono de las vibraciones externas.

Además, y si nos encontramos con una grabación en exteriores, la protección contra el viento se vuelve imprescindible. de otra forma, éste será recogido por las cápsulas y derivando en una toma inservible. Para ello, la opción más recomendada es el uso de cestas anti viento, también conocidas como *wind shield*, tipo *Rycote*® o similar.



Figura 4.22: Rode NT-SF1 con suspensión y anti viento de serie.[40]



Figura 4.23: TSL Soundfield ST450 mkii montado sobre cesta anti viento Cinela Pianissimo.[39]

Posición

En cuanto a la posición son importantes varios factores a la hora de realizar una buena captación del campo sonoro. Uno de los primeros puntos en los que debemos fijarnos es en la orientación del micrófono. Al tratarse de micrófonos capaces de diferenciar entre direcciones, es importante una co-

recta orientación de los mismos. Para ello vienen marcados con un punto y/o marca que nos indica la parte frontal del campo sonoro. Dicha marca se puede apreciar en las figuras 4.24 y 4.25.



Figura 4.24:
Rode
NT-SF1.



Figura 4.25: TSL Soundfield ST450
mkii y su box de control.

Además, la orientación del micrófono puede ser alterada de dos formas distintas. Éstas dos configuraciones se conocen como *endfire* e *invert*. Partiendo de que la posición natural del micrófono éste se encuentra colocado en vertical y con el frontal hacia la parte frontal deseada para nuestro campo sonoro, serán varias las posiciones que podremos utilizar a partir de las opciones *endfire* e *invert*. Dichas opciones se pueden observar en la figura 4.26, donde dependiendo de las necesidades del espacio donde vayamos a grabar, deberemos vernos obligados a una u otra configuración. En esencia, el micrófono puede estar en su posición original, en posición invertida o boca abajo, en posición *endfire* u horizontal, o una suma de estas dos últimas.

Es importante tener este esquema claro. El principal motivo es que en nuestro conversor de formato A a formato B, deberemos indicarle en qué posición se encuentra el micrófono para que la decodificación se lleve a cabo correctamente, con la parte frontal del campo sonoro dónde debería estar y todas sus direcciones espaciales situadas correctamente. Esta indicación siempre se llevará a cabo en la parte de la cadena donde se realice la conversión entre formatos AB, ya sea ésta una conversión por *software* o por *hardware*.

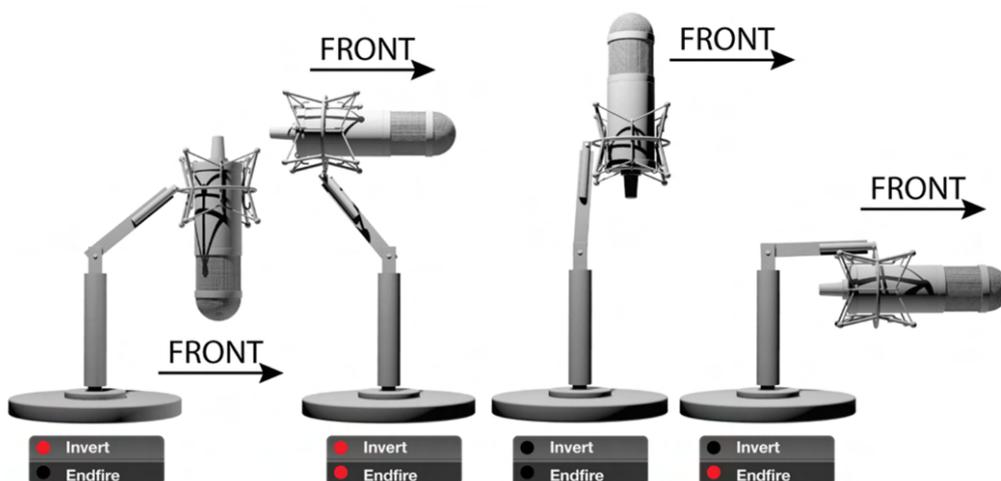


Figura 4.26: Posibles posiciones para un micrófono Ambisonics.[41]

4.2.2 Interfaz propietaria

El siguiente paso en nuestra cadena es opcional y dependerá del micrófono que estemos utilizando. Existen micrófonos Ambisonics que entregan al usuario 4 canales en formato A. De esta forma, la conversión AB se realiza vía *software* ya en nuestro ordenador, por lo que este punto de la cadena de grabación no es necesario.

Sin embargo, son bastantes los micrófonos, especialmente los HOA, que utilizan algoritmos propietarios para la conversión entre formatos, así como interfaces de conexión con el ordenador digitales, ya sea USB, Firewire, etc. Cómo no podemos recopilar aquí todos los micrófonos del mercado con éste tipo de particularidades, a continuación se va a ver tanto el micrófono *ST450* cómo el *em32*, ambos con una interfaz propietaria a modo de puente y disponibles en el laboratorio.

TSL SoundField ST450 mkii

Si hablamos del micrófono *TSL SoundField ST450 mkii*, éste incorpora una de estas interfaces propietarias a modo de puente.

En primer lugar, la caja recibe las 4 señales del micrófono en formato A a través de un cable con conectores tipo *LEMO*® de 12 pines. Desde aquí seleccionaremos la ganancia de las cápsulas, así como la posición del micrófono necesaria para la conversión AB. Trás esto, podremos sacar la señal en formato M/S ó estéreo con el ancho y el patrón polar virtual que deseemos por una salida dedicada, así como monitorizar en formato binaural directamente a través de su salida de auriculares dedicada. Por último, la caja nos sacará 4 señales en formato B ya convertido para grabarlas en el soporte que elijamos a través de dos puertos de 5 pines. El primero de los puertos nos entregará las señales W y X, mientras que el segundo serán Y y Z.

Otra de las funciones que nos ofrece la interfaz del *ST450 mkii* es la de prevenir los efectos negativos de la condensación. Es bastante usual que en micrófonos de condensador polarizados se pueda dar un efecto de condensación de la humedad en la cápsula con los cambios de temperatura. Este efecto es muy negativo en la captura, por lo cuál éste micrófono ofrece una medida para mitigarlo. Se trata de una alimentación de aproximadamente 12V en las cápsulas a modo de calentador, de forma que éstas no sufran cambios repentinos de temperatura que puedan dar lugar a dicho fenómeno.



Figura 4.27: Interfaz propietaria con conversión AB y monitorización del *ST450 mkii*.

Alternativa al *box* del *TSL SoundField ST450 mkii*

Si bien es cierto que el *box* del micrófono *ST450 mkii* hace de él una herramienta versátil, con la posibilidad de monitorizar en tiempo real y en formato binaural nuestra captación, se puede volver algo falto de versatilidad al realizar *in-situ* la conversión AB. Podría ser interesante registrar el material en formato A sin convertir para posibles pruebas con conversiones vía *software*, por ejemplo. Es por ello, que a continuación se ofrece una solución para prescindir del *box* de serie.

La solución pasaría por construir una caja con un conector hembra *LEMO*® de 12 pines, donde

conectaremos el micrófono, y una salida con 4 puertos XLR que contendrán las señales **FLU-FRD-BLD-BRU** de las cápsulas directamente en formato A (SoundField renombra éstas señales como se aprecia en la tabla 4.1). Además, deberemos conectar un puerto jack DC y alimentarlo con un voltaje de 12V DC para el sistema de calentamiento de las cápsulas. De esta manera, obtendremos directamente las señales en crudo para grabarlas en nuestro soporte deseado en formato A, pudiendo realizar la conversión posteriormente vía *software* a través de aplicaciones de terceros. A continuación se expone el esquema de conexiones de los 12 pines según el fabricante:

Pin	Concepto	Señal
1	LB (+)	FRD
2	LB (-)	FRD
3	RB (+)	FLU
4	RB (-)	FLU
5	LF (+)	BRU
6	LF (-)	BRU
7	RF (+)	BLD
8	RF (-)	BLD
9	Voltage GND	
10	-V	
11	+V	
12	Signal GND	

Tabla 4.1: Esquema de conexionado para el conector de 12 pines del *ST450 mkii*

Eigenmike em32

Para el caso del *Eigenmike em32* el procedimiento de conexionado se vuelve bastante diferente. Al tratarse de un micrófono HOA de 32 cápsulas, el conexionado analógico se convertiría en una tarea muy complicada, especialmente al necesitar de 32 pre amplificadores de micrófono para hacerlo funcionar. Para solucionar la situación, el *em32* incorpora 32 PGA⁵, uno para cada cápsula, dentro del propio micrófono, así como sus respectivos conversores AD. De esta manera, la señal se ajusta a la ganancia adecuada y se digitaliza directamente en el propio micrófono, siendo la conexión de salida en formato digital.

Para controlar cada PGA se establece una conexión MIDI entre el micrófono y su *box* propio. De esta forma, desde el PC, que estará conectado también al *box*, controlaremos el micrófono. Las señales saldrán del *em32* en formato digital a través de un flujo de señales propietario. Dichas señales llegarán al *box*, a partir de ahora EMIB (*Eigenmike Interface Box*). En este punto, éstas señales se convertirán en el formato que deseemos, siendo posible formato Ambisonics de hasta cuarto orden, y de ahí a través de una interfaz Firewire o MADI (es necesaria una tarjeta de expansión), al PC.

De esta manera, el procesado de señal se realizará en el EMIB, y no en el PC, quitándole carga computacional. Además, el EMIB funciona como puente de comunicación entre el ordenador y el micrófono, facilitando el control de éste y ofreciendo la posibilidad de comunicación de baja latencia con 32 canales separados gracias a su interfaz FW/MADI.

Por tanto, para el caso el *em32*, la EMIB es algo más que un simple conversor de formatos, y actúa como puente de comunicación ente el PC y el micrófono, controlando a éste último, calibrándolo y

⁵PGA.- Programmable Gain Amplifier.

ofreciendo al PC una interfaz de audio de 32 canales, con conversión de formatos, comunicación y control. Además, cada EMIB acepta hasta 2 *em32* a través de sendos puertos Ethernet.



Figura 4.28: Interfaz propietaria EMIB. (Arriba.- Frontal/Centro.-Trasero/Abajo.-Trasero con interfaz MADI.)

4.2.3 Interfaz de audio

El siguiente paso únicamente es válido para los micrófonos con salida analógica. Esto quiere decir que cualquier micrófono HOA con comunicación digital como *Eigenmike em32*, *Zylia ZM-1* o *Voyage Audio Spatial Mic* no necesitará de una interfaz de audio externa ya que ésta se encuentra incorporada en el propio micrófono, ofreciendo así un todo en uno con pre amplificadores, conversores AD e interfaz de comunicación con el PC.

Por otro lado, existen otro tipo de micrófonos (la mayoría) que ofrecen una salida analógica, entregando señales analógicas de nivel micrófono. Este tipo de micrófonos requieren de una serie de componentes externos para hacerlos útiles. Estos componentes son:

- Un pre amplificador por cápsula, encargado de aumentar el nivel de la señal de nivel de micrófono a nivel de línea con una calidad suficiente como para no introducir distorsión.
- Un conversor AD para cada señal con una buena resolución y alta frecuencia de muestreo.
- Una interfaz de audio que permita la comunicación con un ordenador. Esta parte puede sustituirse por un soporte de grabación independiente del ordenador como una grabadora de campo.

En la figura 4.29 podemos observar un esquema con el flujo de la señal para los bloques que se han explicado. Tratándose de un micrófono FOA como los que se han explicado, por ejemplo el *Rode NT-SF1* del laboratorio, el flujo de la señal sería el siguiente:

- Cada cápsula es conectada mediante un cable XLR a la entrada de un pre amplificador de micrófono. (Los colores siguen el cableado original del *NT-SF1*).
- Tras amplificar cada señal a un nivel de línea, la señal, que todavía es analógica, es convertida a digital mediante un conversor AD dedicado. La profundidad de bits y la frecuencia de muestreo de los conversores será determinante para una buena toma.
- Tras la conversión, son dos los caminos que podemos tomar. En primer lugar podremos registrar cada una de las 4 señales en un soporte de grabación como una tarjeta SD o un HDD a través de una grabadora de campo (existen grabadoras de campo con decodificador Ambisonics a binaural para monitorización en tiempo real). La otra opción es utilizar una interfaz de audio USB, o el protocolo de conexión que deseemos, conectarla a un ordenador y utilizar éste como soporte de grabación.

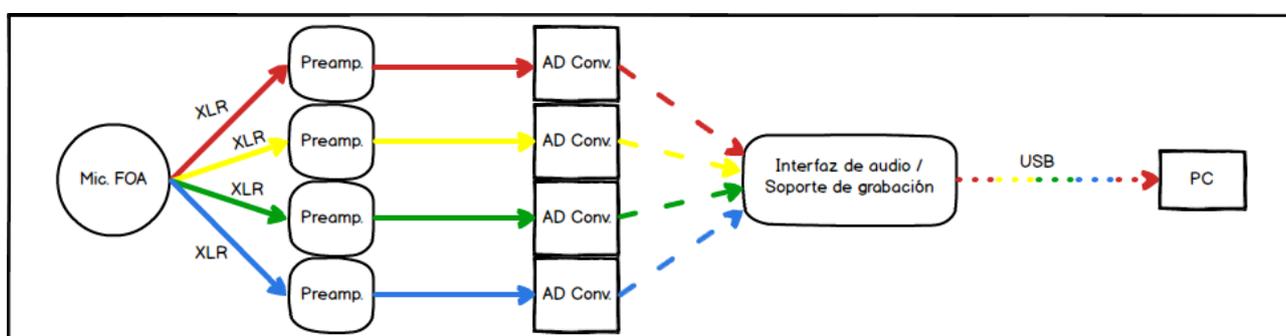


Figura 4.29: Flujo de señal en grabaciones Ambisonics FOA.

Por tanto, para grabaciones FOA, se necesitará una tarjeta de audio o una grabadora de campo con al menos 4 entradas de micrófono. Si utilizásemos un micrófono de orden superior como el *CoreSound OctoMic*, el número de entradas de micrófono ascendería hasta 8. A ser posible, es preferible elegir sistemas donde la ganancia de los pre amplificadores se pueda controlar digitalmente, de manera que sea más sencillo ajustar los 4/8 pre amplificadores a la misma ganancia para una buena toma del campo sonoro. Si esto no fuera posible, se recomienda realizar un ajuste lo más fino posible de las ganancias, ya que sin esto, el campo sonoro se verá alterado y las señales no se capturarán con la misma ganancia.

Como ya se ha dicho anteriormente, lo usual es que para micrófonos HOA, éstos tengan una conexión digital con el ordenador e incorporen su propio sistema de pre amplificadores y conversores para no depender de equipos externos.

4.2.4 PC (Software)

Este punto es variable como ya se a mencionado previamente. Existe la posibilidad de que utilicemos como soporte de grabación una grabadora portátil. La otra opción es utilizar una tarjeta de sonido con interfaz de audio y utilizar el ordenador como soporte de grabación. En cualquier caso, grabaremos en formato A o formato B para el *ST450* y su *box*. Una vez tengamos las 4 pistas pertenecientes al formato A/B, éstas deberán ser procesadas.

Para ello, cada fabricante suele facilitar un *software* con sus micrófonos, de manera que al realizar la conversión AB se apliquen los filtros de compensación para las cápsulas, corrección de la distancia al centro, etc. De esta forma, y utilizando dicho *software*, nos aseguraremos de una conversión óptima

entre formatos. Además, es posible utilizar programas de terceros para dicha conversión, siempre que éstos incluyan *presets* para los micrófonos que estamos utilizando. Es importante remarcar que únicamente uno de esos programas podrá ser utilizado en cada cadena Ambisonics. Estos *software* suelen venir en formatos estándar de trabajo dentro del audio por ordenador, como en formato VST, AU o VST3 por ejemplo.

En las figuras 4.30 y 4.31 observamos los VST de conversión para el *NT-SF1* y para los micrófonos analógicos de *TSL SoundField*. De esta manera, y directamente desde ellos, podremos realizar la conversión AB, así como la selección de la posición del micrófono, incluso la decodificación a diferentes formatos *surround*, mono o estéreo.



Figura 4.30: VST del *Rode NT-SF1*. para manejo y conversión entre formatos. Incluye decodificación.

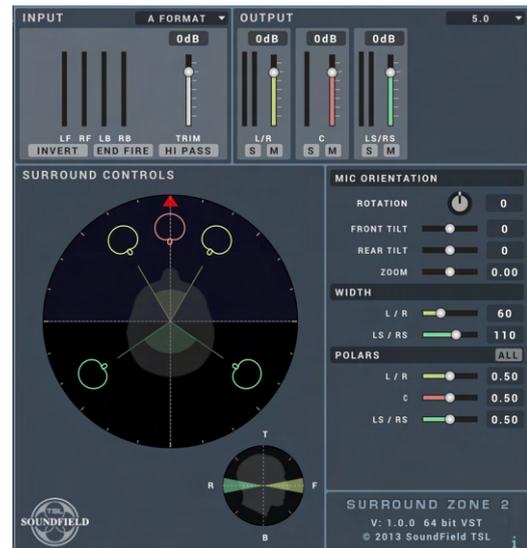


Figura 4.31: VST *Surround Zone 2* de *TSL Soundfield* para manejo de sus micrófonos analógicos.

En el VST de *Rode* se aprecia cómo en la entrada se ha seleccionado el *NT-SF1*, y para la salida el formato AmbiX. Por el contrario, en el *Surround Zone 2* de *TSL SoundField*, se ha elegido una entrada en formato A, por ejemplo utilizando la alternativa expuesta en la sección 4.2.2, y para la salida se busca decodificar el campo sonoro en un sistema de reproducción 5.0. Además, se pueden modificar varios parámetros como el patrón polar virtual de las fuentes, la orientación de éstas frente al oyente, el ancho, la elevación de las fuentes o la mezcla de fuentes.

Si hablamos del *software* de control y procesamiento del *Eigenmike em32*, la cosa se vuelve diferente. Al tratarse de un micrófono HOA, donde el control del mismo se hace de forma digital, y donde su interfaz es propietaria, éste viene con un completo paquete *software* de varios componentes para trabajar con él. A continuación se expone un breve resumen de las herramientas facilitadas por le fabricante, así como su propósito y características más destacadas:

- ***em32setGain.***- Se trata de un *software* autónomo para control del *em32*. Su funcionamiento es simple: consiste en un programa que se comunica con el micrófono vía mensajes MIDI. Mediante esta comunicación, se puede establecer la ganancia de los PGA del micrófono, activar la calibración de las cápsulas (la cuál está guardada de fábrica en una memoria no volátil dentro del micrófono), así como verificar el estado del micrófono y la comunicación de éste con el ordenador. Está pensado para su uso con algún DAW de terceros, pudiendo así controlar el micrófono desde una aplicación autónoma y garantizando el buen funcionamiento del micrófono. Se pueden

observar mensajes de verificación para la conexión con el EMIB, con el propio micrófono, con la interfaz MADI, así como verificar la calibración del micrófono. Además, nos muestra los valores de calibración de fábrica para cada una de las cápsulas, tal y como se observa en las figuras 4.32 y 4.33.

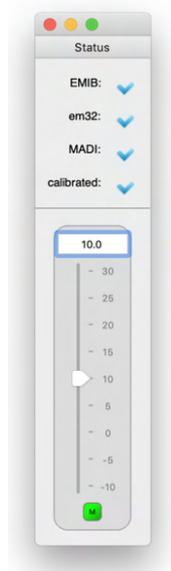


Figura 4.32:
Interfaz
del VST
em32setGain. [42]

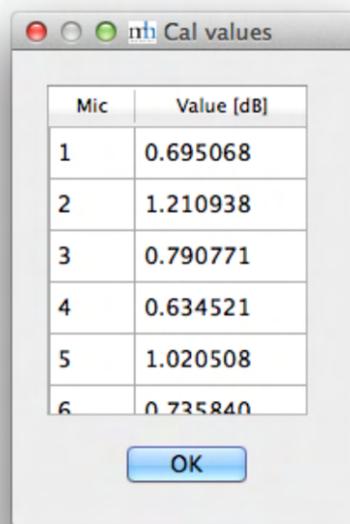


Figura 4.33: Ventana con la calibración
de las cápsulas del *em32*. [42]

- **EigenUnits®.**- Para trabajar desde un DAW y tener control del micrófono, así como de conversión entre formatos o *beamforming* tenemos la *suite* de plugins EigenUnits® en formato VST. Está compuesta por 3 VST independientes para insertar en la pista de nuestro DAW y poder trabajar con el *em32*. A continuación un breve resumen de cada uno:

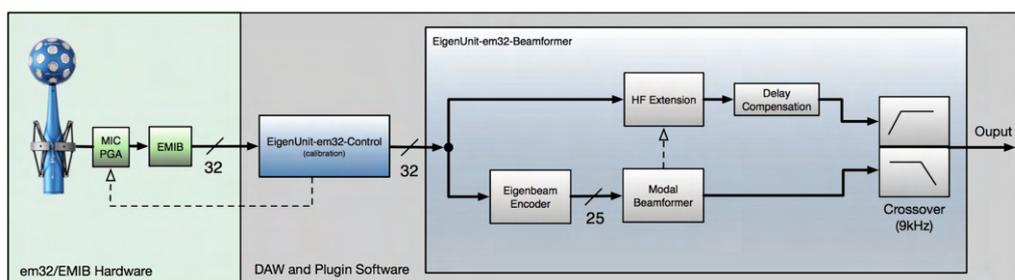


Figura 4.34: Posible diagrama de bloques para trabajar con el *em32* y las herramientas *EigenUnits®*. [43]

- **Control.**- Se trata de una herramienta similar al *em32setGain* pero en formato VST. De esta manera, podremos incorporarla a la pista deseada de nuestro DAW y controlar la ganancia, calibración y comunicaciones del micrófono. Es importante tener en cuenta que únicamente debemos utilizar una de estas herramientas (*em32setGain* o *EigenUnits Control*), ya que el control del micrófono y la comunicación debe hacerse por un único canal.

- **Encoder.**– En este caso, el VST actúa como un *encoder* Ambisonics normal. Recibe 32 entradas calibradas desde el plugin de control (o desde un archivo), y codifica en formato Ambisonics de hasta cuarto orden, con el orden de canales y la normalización que deseemos. Es importante trabajar siempre con buses de mínimo 32 canales. También es importante tener en cuenta que debido a la naturaleza física de Ambisonics, a mayor orden, los armónicos esféricos de dicho orden sufrirán un filtrado paso alto mayor. Concretamente los cinco armónicos de segundo orden están filtrados a 400Hz, 1kHz para los siete armónicos de tercer orden, y 1.8kHz para los nueve de cuarto orden.
- **Beamformer.**– El VST *Beamformer* recibe 32 señales calibradas desde el plugin de control (o desde un archivo), y produce una salida en forma de haz de un sólo canal según los parámetros elegidos por el usuario. Se pueden elegir desde patrones virtuales omnidireccionales, cardioides, super e hipercardioides o figura de 8. Se pueden elegir órdenes de patrón de 0 a 4, así como dirigir la dirección del haz en un espacio tridimensional completo. De ésta forma, con el *em32* podemos simular casi cualquier tipo de micrófono en lo que a respuesta polar se refiere.

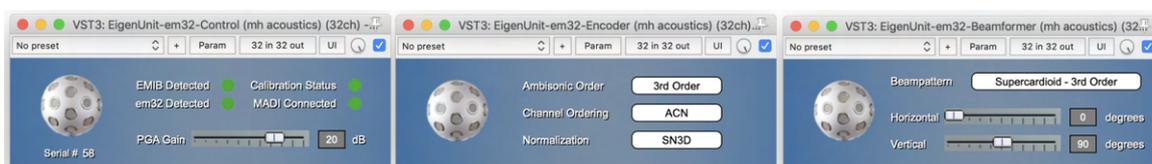


Figura 4.35: *EigenUnits*® en formato VST en el DAW *Reaper*.^[43]

- **EigenStudio**®.- Se trata de un programa *standalone*⁶ que permite la grabación, procesamiento y control del *em32*. El objetivo principal de este programa es cubrir los 3 ámbitos siguientes:
 - Ofrecer una amplia gama y tipos de *beamforming*, tal y como hace el VST *Beamformer* del paquete *EigenUnits*®.
 - Cubrir la codificación de los armónicos esféricos de Ambisonics para HOA hasta cuarto orden, tal y como hace el VST *Encoder* del paquete *EigenUnits*®.
 - Control, realización y edición básica de grabaciones en bruto con el *em32*.

Además, permite renderizar hasta un total de 30 canales de salida distintos con forma de haz simultáneamente (*beamforming*). Permite la selección de múltiples *beampatterns*, o patrones de haz preestablecidos, ofreciendo así una versatilidad enorme. Cada haz virtual puede ser posicionado de manera independiente en cualquier dirección dentro de un campo sonoro 3D, lo que viene a simular casi cualquier micrófono, pudiendo situarlo dónde el usuario desee dentro del espacio sonoro y en post producción.

Por otro lado, los canales de entrada, hasta un máximo de 32, pueden ser seleccionados o bien de señales provenientes de un micrófono *em32* en tiempo real, de un archivo proveniente de una grabación con el *em32*, así como de material Ambisonics previamente codificado del *em32*. Los canales de salida podrán ser reproducidos a través una salida física de algún tipo de controlador de audio, así como ser guardados en un archivo para su almacenamiento.

Por tanto, *EigenStudio*® ofrece una solución total para el manejo del micrófono *em32* en todas sus vertientes, además de prescindir de software de terceros para su uso. De esta forma, podemos controlar, grabar, editar y codificar el material del *em32* en formato Ambisonics o mediante la técnica de *beamforming* con una sola herramienta y sin necesidad de extras.

⁶*Standalone.*– Que funciona sin necesitar de ningún programa externo, cómo un VST dentro de un DAW.

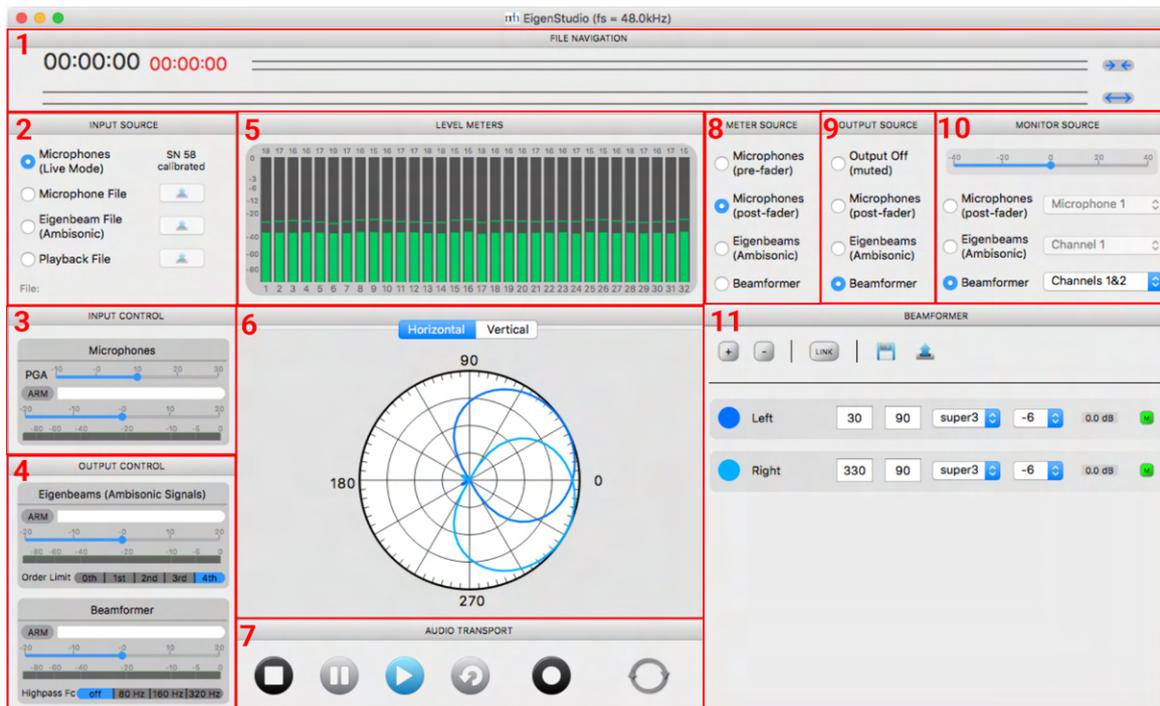


Figura 4.36: Ventana principal del software independiente *EigenStudio*®.[44]

Ahora se va a ver brevemente cada una de las secciones de las que se compone el *EigenStudio*® con el fin de entender cómo funciona y qué posibilidades nos ofrece este programa autónomo.

1. La sección “*File Navigation*” muestra la barra de navegación de archivos encargada de proporcionar al usuario información gráfica sobre la posición actual en el archivo de audio durante la reproducción. Permite al usuario poder seleccionar secciones de audio específicas, así como poder moverse entre posiciones y más.
2. La sección “*Input Source*” permite al usuario elegir la fuente de audio con la que desea trabajar. Las opciones disponibles son: *Microphone (Live Mode)*, para trabajar con un *em32* en modo directo, *Microphone File* para trabajar con una grabación anterior del *em32* sin procesar, *Eigenbeam File (Ambisonics)* para trabajar con un archivo de audio del *em32* previamente codificado en formato Ambisonics de hasta cuarto orden y por último *Playback File* para trabajar con un archivo de audio previamente grabado únicamente para su escucha, con todo el procesamiento deshabilitado.
3. La sección “*Input Control*” en primer lugar permite el ajuste de la ganancia de los PGA. A continuación permite armar el micrófono para realizar una nueva grabación, indicando el nombre del archivo generado. Por último, permite ajustar el volumen (cómo si de un *fader* se tratara) y muestra un vúmetro con el nivel de grabación.
4. La sección “*Output Control*” permite al usuario elegir qué señales serán las elegidas para cada salida física de audio. Al trabajar con el EMIB, éstas pueden ser cualquiera de los 8 canales ADAT que incorpora. Por otro lado, se pueden elegir como salidas físicas cualquiera que incluya nuestra interfaz de audio, ya sea interna del ordenador o externa. Siempre será importante tener en cuenta el número de salidas físicas de nuestro controlador ya que si el número de salidas físicas es inferior al número de salidas virtuales de nuestro proyecto Ambisonics, únicamente sonarán los M primeros canales. Para escucha binaural con una salida estéreo de auriculares será suficiente.

5. La sección “*Level Meters*” muestra los vúmetros con los niveles de cada cápsula. Los números de la parte inferior indican el número de canal/cápsula del *em32*. La escala está en dBFS (con 0dB de ganancia en el PGA, 0dBFS corresponden aproximadamente con 124dB SPL). La barra sólida representa el nivel RMS mientras que la línea delgada muestra el valor de pico. Los indicadores se vuelven de color rojo cuando el nivel supera los -9dBFS. El número que aparece en la parte posterior muestra el valor máximo de pico que experimentó el canal correspondiente. Éste valor de pico se puede resetear haciendo doble clic.
6. La sección donde se muestran los patrones polares muestra, de manera visual, los patrones polares virtuales creados en la sección 11 mediante la técnica de *beamforming*. Muestra una vista Horizontal y Vertical.
7. La sección “*Audio Transport*” muestra la barra de transporte para movernos por el archivo de audio, reproducirlo, pausarlo, activar la grabación o activar el modo *loop*.
8. La sección “*Meter Source*” permite mostrar diferentes valores en la sección número 5. La primera opción permite mostrar la ganancia de los 32 canales del *em32* antes de la ganancia aplicada en la sección *Input Control*. La segunda es similar pero muestra los valores después de aplicar el volumen de la sección 3. La tercera nos muestra los valores de los canales después de la codificación en formato Ambisonics. Por último, nos muestra los valores de los canales tras la codificación con la técnica de *beamforming*.
9. La sección “*Output Source*” permite al usuario elegir qué señales desea encaminar a las salidas físicas de su *hardware* de audio. Al trabajar con la EMIB del *em32*, las salidas físicas pueden ser la salida ADAT de 8 canales. De otro modo, el usuario podrá utilizar otro *hardware* de audio como salida física, siempre atendiendo al número de salidas físicas disponibles y al número de salidas virtuales.
10. La sección “*Monitor Source*” permite al usuario monitorizar cualquiera de las señales de entrada o salida. Par ello, se hace uso de la salida de auriculares del EMIB, o de otra salida física de nuestra interfaz de audio. El usuario podrá ajustar volumen de salida de monitoraje o seleccionar cualquiera de las señales dentro de la cadena de audio del programa para su monitorización.
11. La sección “*Beamformer*” se trata de la sección principal para la funcionalidad de *beamforming* del *em32*. Desde aquí, el usuario puede definir el número total de salidas *beamforming*, seleccionar y dirigir cada haz en cualquier ángulo deseado, así como establecer ganancias independientes para cada uno de ellos. Se pueden añadir hasta 30 canales *beamforming*, editar cada uno de ellos, guardar configuraciones preferidas por el usuario o ajustar la posición y la ganancia tal y como se ha mencionado. Además, cada haz puede ser visualizado mediante su patrón polar en la sección 6 de la figura 4.36.

Capítulo 5

Mezcla en Ambisonics

En el capítulo anterior se han visto posibles sistemas de reproducción para material Ambisonics, tanto a nivel doméstico en formato binaural, como a niveles más profesionales con sistemas multicanal de gran formato. El sistema final que utilice el usuario es indiferente siempre que haga uso de un buen diseño y un decodificador bien implementado, tal y como se han visto las distintas opciones y recomendaciones existentes.

Por otro lado, se ha hecho un repaso sobre los métodos y la forma de trabajo para realizar grabaciones directamente en formato Ambisonics a través de los diferentes micrófonos de los que dispone el departamento, si bien éste repaso se ha visto de forma teórica al no poder realizarse de forma práctica por limitaciones debidas al COVID-19.

Por tanto, y teniendo claro cómo grabar y cómo reproducir material Ambisonics, ha llegado el momento de ver el paso intermedio, es decir, la mezcla y la postproducción de material Ambisonics. Para ello se van a explicar tres ejemplos desde diferentes enfoques para poder ver las diferentes formas de trabajar con esta tecnología:

- **1. Material inicial: grabación Ambisonics y tomas *spot*.** En este ejemplo se partirá de un material grabado originalmente en formato Ambisonics de orden 1, así como de varias tomas tipo *spot* o tomas de instrumento (tomas de un micrófono monofónico colocado en cada instrumento de forma individual). Dicho material será externo a nosotros, por lo que únicamente tendremos el control sobre la mezcla y el procesado del mismo. La idea es poder ver cómo trabajar con el, y que resultado final podemos llegar a alcanzar. Este ejemplo será el más representativo en cuánto al flujo de trabajo con música en formato Ambisonics. La idea es poder mostrar cómo se trabajaría para una mezcla musical en formato Ambisonics obteniendo un formato de escucha binaural adecuado para disfrutar de música con los atributos y posibilidades espaciales que caracterizan al Ambisonics. La mezcla final se realizará en Ambisonics de orden 3.
- **2. Material inicial: grabación multipista.** En este ejemplo partiremos de una grabación multipista estándar descargada de internet. Se realizará una mezcla en formato Ambisonics de orden 3 para reproducción en formato binaural, así como una mezcla en formato estéreo y se realizará un breve comparación de resultados.
- **3. Material inicial: multipista sintético.** En este ejemplo trabajaremos con una producción propia en formato de síntesis, es decir, todo realizado desde el propio ordenador. De esta forma, buscaremos realizar unos movimientos espaciales de los sonidos de una forma agresiva, buscando exprimir y acentuar las virtudes y posibilidades que nos ofrece Ambisonics, a pesar de sacrificar

“naturalidad” en la escucha. En este ejemplo se codificará el material en orden 3 y en orden 5 para realizar una breve comparación de resultados mediante monitorización binaural.

5.1 Caso práctico 1: Ambisonics + *spot*

Para el primer caso práctico vamos a hacer uso de un material disponible a través de la escuela *Berklee Online*, en su artículo *Learn Ambisonics with John Escobar* [45]. Se trata de la obra “*Crisantemi*” del compositor *Giacomo Puccini*, grabada en el *Boston Symphony Hall*. A través de este artículo se nos facilita la descarga de material mixto Ambisonics / monofónico para pruebas y aprendizaje. La grabación se trata de un cuarteto de cuerda frotada compuesta por dos violines, un violonchelo y una viola. Además de los 4 micrófonos *spot* de cada instrumento, la grabación se complementa con dos micrófonos monofónicos en posición L/R para capturar parte del estéreo, un micrófono Ambisonics *Sennheiser*® *AMBE0* de primer orden de audiencia/ambiente, y un micrófono Ambisonics del mismo modelo de primer orden principal o conductor.

El proyecto de *Reaper* está configurado a una resolución de **24 bits** y una frecuencia de muestreo de **48 kHz**.

El diagrama de situación de los micrófonos en el espacio se ha obtenido a partir del vídeo facilitado en el artículo y es aproximado, por lo que las distancias son estimaciones. En la figura 5.1 podemos observar la situación de cada micrófono dentro del espacio. Es importante destacar éste aspecto ya que, al ser una mezcla multi micrófono, y no poder ajustar los retardos entre micrófonos de forma exacta, lo más probable es que tengamos problemas de fase al sumar varios micrófonos. Al tratarse de un ejemplo práctico, este factor no se tendrá demasiado en cuenta, pero viene bien tenerlo presente en caso de realizar tomas multi micrófono.

Se ha tomado el micrófono Ambisonics principal o conductor como el punto de escucha del oyente, por lo que tanto a los micrófonos L/R como al micrófono de ambiente, se ha calculado el *delay* necesario para compensar la distancia al oyente (siempre aproximado), tal y como ya se ha mencionado. Para el cálculo del *delay* se ha tomado una velocidad del sonido estándar de 340 m/s.

Por tanto, la posición de los micrófonos, la compensación de la distancia, así cómo el orden y los colores de la sesión en *Reaper* se pueden ver en el diagrama de la figura 5.1.

A partir del material explicado, a continuación se va a hacer un repaso al flujo de audio y el procesamiento del mismo que se ha llevado a cabo en *Reaper*.

Cómo se observa en la figura 5.2, el número de *plugins* empleado no es muy elevado, salvando el bus general Ambisonics, donde se ha realizado una especie de “masterización” general para darle a la mezcla el color y la energía necesaria. **Es importante destacar que en todos los casos se tratan de pistas con 16 canales para trabajar en Ambisonics de orden 3.**

Como puntos importantes a tener en cuenta, y los cuáles serán generales para trabajar en Ambisonics:

- 1. El pan/balance por canal se dejará centrado. Es importante no tocar el pan para no alterar la posición de la fuente en el campo sonoro. Para la situación espacial utilizaremos únicamente los codificadores/*panners* Ambisonics.
- 2. Los envíos los haremos en formato multicanal y enviaremos los 16 canales de cada pista al bus general.

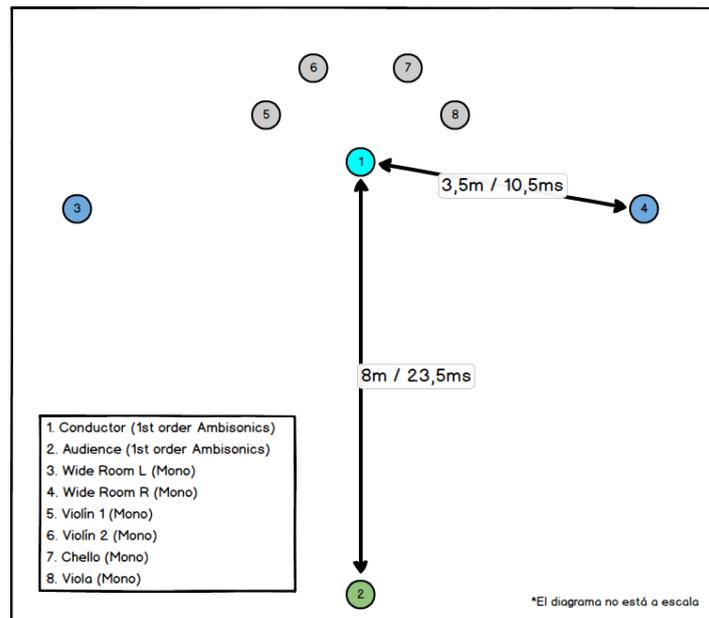


Figura 5.1: Diagrama de situación de los micrófonos con los tiempos de *delay* necesarios y la lista de canales correspondiente.

- 3. Desconectaremos la salida por el máster de cada canal independiente. Únicamente sacaremos por el máster el bus Ambisonics general.

A continuación se va a realizar un repaso pista por pista de los efectos insertados y el objetivo de los mismos.

- **1. Conductor.-** Se trata del micrófono Ambisonics principal, situado en el punto de escucha virtual del oyente. Por ello, el único efecto que se ha utilizado es el *Compass upmixer* del paquete SPARTA para elevar de primer a tercer orden, ganando así matices a nivel espacial.
- **2. Audience.-** Micrófono Ambisonics de primer orden situado en la zona de audiencia. Para éste, se ha elevado el material a tercer orden. A continuación se ha utilizado el *plugin Room Encoder*, en la figura 5.3 del paquete IEM para situar la fuente alejada, añadiendo las reflexiones naturales de la sala y la coloración de la misma o los retardos necesarios.
- **3-4. Wide Room L/R.-** El procesado de los canales es similar ya que ambos son simétricos. Para ello se han codificado con el *plugin ambiENC* del paquete SPARTA en tercer orden. Cada uno de ellos se han situado a $100^\circ/-100^\circ$. A continuación se les ha aplicado un *delay* simple para compensar la distancia de éstos hasta el punto del oyente.
- **5-8. Instrumentos.-** Para los canales independientes de cada instrumento únicamente se ha utilizado el *ambiENC* para colocar cada instrumento dentro del espacio en su posición y codificarlo en tercer orden Ambisonics. Para los instrumentos se ha decidido no añadir un *delay* para compensar la distancia al estar éstos a 1 metro del micrófono principal, de forma que así ahorramos procesado y simplificamos el proceso. Además, se hizo una prueba y el resultado entre aplicar o no *delay* era inapreciable a nivel auditivo o de problemas de fase, por lo que se decidió prescindir de él.

Cada uno de estos 8 canales se envía a un bus general llamado AMBISONICS, en color rojo. En este bus se realiza el procesado general de la mezcla, se decodifica en formato binaural y se saca por



Figura 5.2: Mezclador con los *plugins* insertados, los colores del diagrama y el bus *Ambisonics* general.

el máster para monitorización por auriculares. A continuación se va a ver detenidamente el procesado en el bus **AMBISONICS**:

- **1. IEM MultiEQ.**- En primer lugar se ha insertado el *MultiEQ* del paquete IEM (ecualizador paramétrico Ambisonics). En él únicamente se ha colocado un paso altos para limpiar la zona útil del espectro (al ser música acústica en esta zona no tenemos nada de información, únicamente ruido). Además, se ha suavizado la zona de los medios-graves y se ha enfatizado la zona de los 5kHz para añadir brillo a la mezcla final. Cómo se ve, se trata de una EQ muy sutil únicamente para realzar un poco el carácter general y levantar las altas frecuencias perdidas durante el procesado Ambisonics.
- **2. IEM MultiBandCompressor.**- Compresor multibanda Ambisonics. En este proyecto se ha utilizado de una forma muy sutil para realzar y dar matices en bandas de frecuencia concretas, aportando brillo y reduciendo la zona de bajas frecuencias, trabajando en colaboración con el equalizador para aportar color y carácter a la mezcla. Posee 4 bandas de compresión independientes en formato Ambisonics sin alterar el campo sonoro.
- **3. IEM FDNReverb.**- Reverb Ambisonics. Se ha colocado en la mezcla una reverb a 4 segundos para situar el pasaje musical en un espacio real, similar a un auditorio para conciertos. También se ha alterado el ecualizador de la reverb para limitar las reflexiones en altas frecuencias, buscando así una reverb más cercana a la realidad, donde las altas frecuencias se atenúan antes que las graves.
- **4. IEM EnergyVisualizer.**- Visualizador de energía para monitorización Ambisonics. Muestra la energía dependiendo de su dirección para observar cómo se comporta el campo sonoro en tiempo real.
- **5. SPARTA AmbiBIN.**- Descodificador Ambisonics a formato binaural. Se utiliza para pasar de formato Ambisonics de tercer orden a escucha binaural. Es importante definir bien el formato de Ambisonics ACN y SN3D. Se utiliza la HRTF por defecto.



Figura 5.3: *Room Encoder* del paquete IEM para codificación de una fuente Ambisonics dentro una sala.

- **6. Cockos ReaComp.**- En la salida binaural se sitúa este compresor nativo de *Reaper*. Únicamente se utiliza como limitador para controlar la salida, por eso el ratio está situado en infinito.



Figura 5.4: Vista general con todos los *plugins* aplicados en el bus AMBISONICS.

Como ya se ha visto, se trata de un flujo de trabajo bastante sencillo. Tratándose de una pieza de música clásica, es importante no abusar del procesado ni de alterar significativamente el sonido natural de la pieza. El procesado a nivel de canal individual es nulo, únicamente habiendo utilizado un codificador Ambisonics por canal para situarlo en el espacio sonoro ó, a lo sumo, un *delay* para compensar la distancia entre fuentes.

Por otro lado, y el único punto de la mezcla donde se ha trabajado más a fondo, es en el bus general Ambisonics. De esta forma, y a modo de “masterización”, se ha buscado realzar parte del timbre y el color perdidos durante la mezcla, especialmente buscando algo más de “brillo”. Este problema con las altas frecuencias es intrínseco al formato Ambisonics y es importante que lo tengamos presente a la hora de trabajar con él.

El resultado se puede escuchar en la pista “Ambisonics - Ejemplo 1 (binaural)” de la carpeta “FINAL MIXES”, adjunta a este proyecto, en formato **.mp3 24 bits - 44.1 kHz**.

5.2 Caso práctico 2: Multipista acústico

En el siguiente ejemplo se va a partir de material multipista facilitado por *Telefunken Elektroakustik* [46]. Se trata de la canción “*Corrina, Corrina*” con todas sus pistas e interpretada por la “*Abletones Big Band*” y grabada en estudio.

A continuación se expone una tabla con la microfonomía y los preamplificadores utilizados para la grabación de cada instrumento. Se ponen a modo de curiosidad para situar cada instrumento en el contexto en el que fue grabado, si bien para el proceso que nos acontece será de poca relevancia.

Intrumento	Micrófono / DI	Preamplificador
Bombo	Telefunken U48	Neve 33102
Ambiente batería	Telefunken AR-70	Neve 33102
Contrabajo DI	DI Telefunken TDA-1	Neve 33102
Contrabajo mic.	Telefunken ELA M 260	Neve 33102
Guitarra	Telefunken AR-51	Presonus M80
Piano L	Telefunken ELA M 260	Neve 33102
Piano R	Telefunken ELA M 260	Neve 33102
Saxo A	Telefunken AK-47 mkii	Neve 33102
Saxo B	Telefunken AK-47 mkii	Neve 33102
Trombón A	Telefunken AR-51	Neve 33102
Trombón B	Telefunken AR-51	Neve 33102
Trompeta A	Telefunken AR-51	Neve 33102
Trompeta B	Telefunken AR-51	Neve 33102
Voz	Telefunken U-48	Neve 33102
Room A	Beyer Dynamic M160	Neve 33102
Room B	Beyer Dynamic M160	Neve 33102
Room C	Telefunken AR-51	Chandler Germanium
Room D	Telefunken AR-51	Chandler Germanium
Ambiente L/R	Telefunken AR-70	Neve 33102

En este caso práctico existen dos proyectos bien diferenciados. En primer lugar se va a realizar una mezcla en formato Ambisonics de orden 3. Por otro lado se va a realizar una segunda mezcla en formato estéreo. El objetivo es disponer de ambas mezclas partiendo de un mismo punto para evidenciar las diferencias existentes entre mezcla estéreo y mezcla de audio espacial, así como para evaluar posibles resultados u opiniones sobre ambas.

Ambos proyectos de *Reaper* están configurados a una resolución de **24 bits** y una frecuencia de muestreo de **48 kHz**.

Para llevar a cabo la explicación de ambos casos, en primer lugar se va a ver la mezcla Ambisonics, explicando detalles de la mezcla y posibles decisiones tomadas. A continuación se verá un rápido vistazo sobre los aspectos importantes de la mezcla estéreo por carecer de interés directo con el presente proyecto.

5.2.1 Caso práctico 2.1 - Proyecto Ambisonics

En este apartado se va a hacer un repaso por el proceso de mezcla del proyecto de *Reaper*. Antes de empezar se van a aclarar unos cuantos puntos importantes del proceso de mezcla, así como una serie de criterios que se han tenido en cuenta a la hora de realizarla.

- **Profundidad de la mezcla.** Para el proyecto se ha decidido realizar una mezcla lo más transparente y superficial. No se ha incluido ningún tipo de procesamiento por canal más allá del codificador Ambisonis. Únicamente se ha alterado el sonido original de la señal en el bus principal a modo de “masterización”.
- **Ubicación de los instrumentos.** Para la localización espacial de los instrumentos se ha seguido un patrón estándar para *big band*. Al no conocer con exactitud la ubicación original de cada instrumento, se ha tomado este criterio por buscar un punto de compromiso. Esta localización puede dar problemas con los micrófonos de ambiente, si bien se ha obviado este punto.

Una vez vistos estos puntos, lo siguiente a repasar es el flujo de señal interno dentro del proyecto. En él se verá el ruteo de la señal entre los diferentes buses de señal y el motivo de este ruteo.

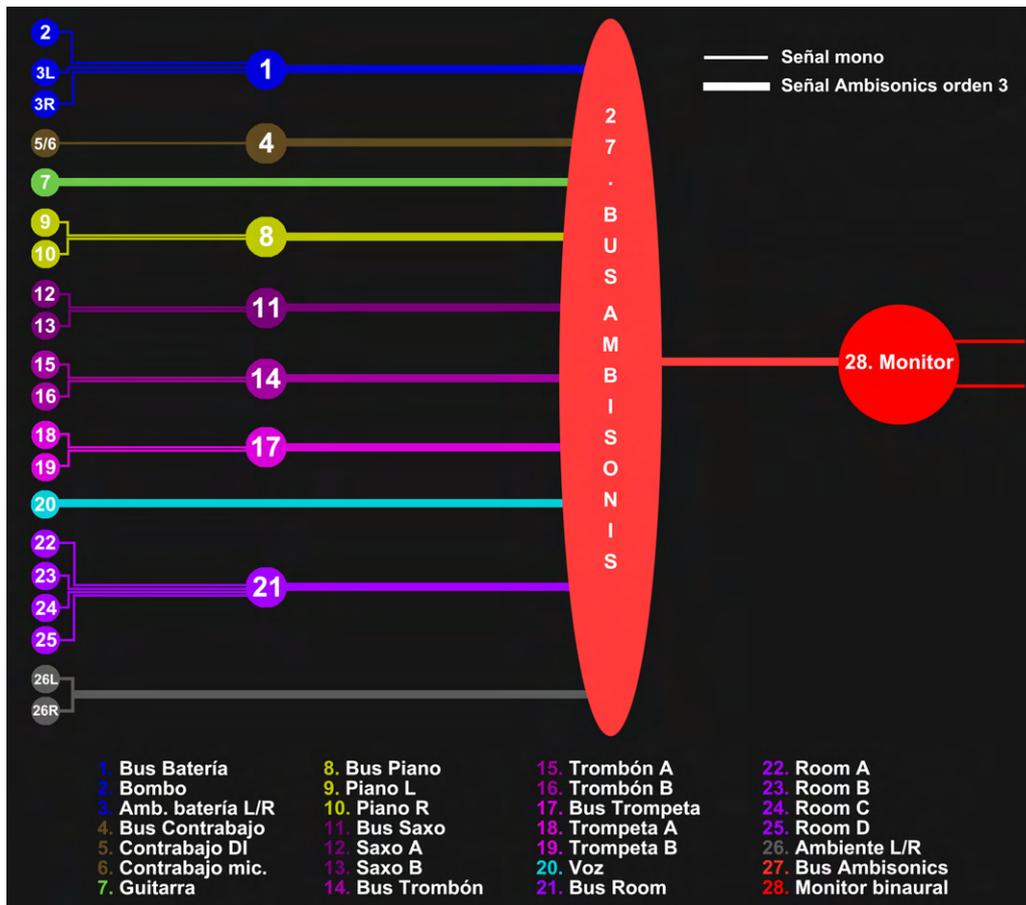


Figura 5.5: Ruteo de la señal entre los diferentes buses y lista de canales.

En primer lugar se va a hacer un repaso a los símbolos y su significado en la figura 5.5:

- **Círculos pequeños:** Los círculos pequeños corresponden a pistas de instrumentos. Éstas pueden ser mono o estéreo.

- **Círculos medianos:** Se trata de buses auxiliares. Se utilizan como paso intermedio para reunir instrumentos de la misma familia y tener un control unificado. También se utilizan para colocar en ellos el codificador Ambisonics y ubicar cada instrumento dentro del campo sonoro.
- **27. AMBISONICS:** Se trata del bus general. A él se envían todas las señales, tanto las provenientes de buses auxiliares como las individuales que no necesitan bus (7. Guitarra, 20. Voz y 26L/26R. Ambiente sala). En éstas últimas, el codificador Ambisonics está incluido en la propia pista de instrumento.
- **28. Monitor:** Se trata del último paso de la señal. Recibe los 16 canales provenientes de la mezcla final hecha en el bus AMBISONICS y la decodifica a formato binaural. De ahí, se envía a las salidas físicas 1/2 de nuestra interfaz de audio.
- **Líneas finas:** Indican, tal y como se ve en la leyenda, señales monofónicas (1 canal).
- **Líneas gruesas:** Son señales Ambisonics de orden 3, es decir, 16 canales.

Por tanto, se puede deducir el proceso de mezcla en cascada. Para cada familia de instrumentos común, se deriva su señal a un bus auxiliar intermedio donde cada una de las señales se codifica dentro del campo sonoro. De ahí obtenemos la señal Ambisonics en orden 3 de esa familia y la mandamos al bus general.

Existen excepciones o casos individuales que es importante ver, ya que estas reglas no son del todo acertadas en todos los casos:

- **3L/3R. Ambiente de batería:** Los micrófonos de ambiente para la batería son una pista estéreo de dos canales. Esos dos canales son enviados junto al bombo (2) al bus de batería, cada uno asignado a un canal del envío. De esta forma, cada canal puede ser codificado de forma individual dentro del codificador.
- **5/6. Contrabajo:** Tal y como se observa, el contrabajo está en las pistas 5 y 6 (no es una pista estéreo). La pista 5 corresponde al contrabajo grabado por una caja de inyección o DI. Por contra, la 6 se trata de una grabación con micrófono. Al tratarse un instrumento de bajas frecuencias, lo usual es situarlo en formato monofónico, por ello ambas pistas se han representado juntas, ya que a pesar de tratarse de pistas separadas (mono), su ubicación en el campo sonoro es la misma.
- **7. Guitarra y 20. Voz:** Tal y como ya se ha mencionado estas dos pistas son un caso particular. A pesar de tratarse de pistas individuales como el resto, éstas no tienen otras pistas de la misma familia. Por ello, se ha evitado incluir buses auxiliares en ambas y se ha procedido a colocar el codificador directamente en las propias pistas. Por tanto, el envío se hace directamente al bus general en formato de 16 canales (orden 3).
- **26L/26R. Ambiente sala:** Por último tenemos el ambiente de sala. Se trata, al igual que el ambiente de batería, de una pista estéreo. La diferencia es que ésta no tiene otras pistas con las que ir agrupada. Se ha situado el codificador directamente a la pista y se ha enviado directamente al bus general, prescindiendo una vez más de un bus auxiliar.

Una vez visto cómo se han ordenado las señales dentro del proyecto, a continuación se va a mostrar un diagrama con la situación de cada instrumento dentro del campo sonoro, figura 5.6. Al igual que en la figura 5.5, en este diagrama se han utilizado los mismos colores para cada pista. Además, estos

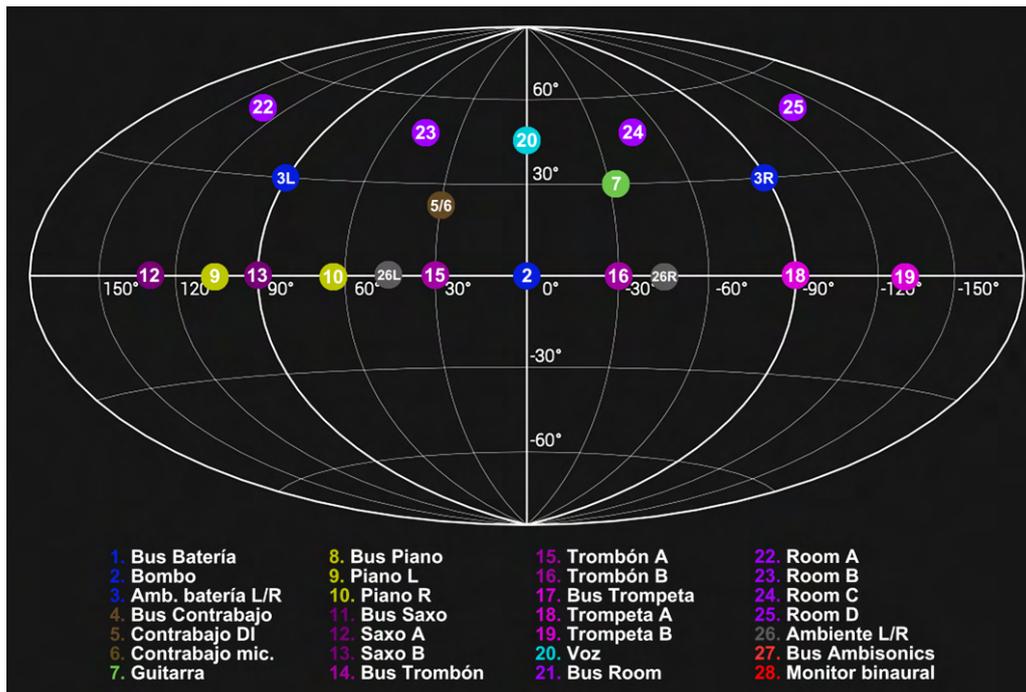


Figura 5.6: Diagrama con la ubicación espacial de cada instrumento y lista de canales.

colores corresponden con los colores originales dentro del proyecto, de forma que sea más fácil situarse dentro del mismo.

Cómo bien se ha dicho, los instrumentos han sido situados en el espacio según la colocación estándar de una *big band* típica. Aún con todo ello, se ha realizado una localización bastante agresiva y abierta, buscando dar la sensación de estar escuchando la banda desde dentro. No se han colocado instrumentos en el hemisferio inferior del campo sonoro por perder gran parte del timbre de los mismo y por ser una posición difícilmente recreable en la realidad.

Una vez todas las señales son ubicadas dentro del campo sonoro, y teniendo todas ellas ajustadas de volumen general, se envían al bus general 27. AMBISONICS. En este bus es donde se realizan el único procesamiento propiamente dicho de la señal. A continuación se hace un breve repaso a los efectos utilizados y su objetivo:

- **1. IEM MultiEQ:** Al igual que en el caso práctico 1, con esta EQ únicamente se busca controlar un poco la zona de bajas frecuencias, donde se ensucia especialmente la mezcla al sumar tantos instrumentos, y agregar un poco de brillo perdido durante la codificación Ambisonics.
- **2. IEM MultiBandCompressor:** Se utiliza únicamente para realzar la banda alrededor de 1 kHz y se controla la zona de bajas y medias-bajas frecuencias realizando una compresión muy sutil.
- **3. IEM FDNReverb:** Se aplica una reverberación larga de unos 5 segundos con una relación señal procesada/señal directa del 40 % para añadir riqueza tímbrica a la mezcla y situarla en un espacio natural.
- **4. IEM EnergyVisualizer:** Visualizador del campo sonoro Ambisonics.

La mezcla general procesada en el bus Ambisonics es enviada por un envío de 16 canales a un bus final de monitorización donde se ha insertado el SPARTA AmbiBIN para decodificar de formato B de

tercer orden a escucha binaural y se ha situado un compresor en modo limitador al final de la cadena para controlar la salida por auriculares.

El resultado se puede escuchar en la pista “Ambisonics - Ejemplo 2.1 (binaural)” de la carpeta “FINAL MIXES”, adjunta a este proyecto, en formato **.wav 24 bits - 44.1 kHz** o en formato **.mp3**.

5.2.2 Caso práctico 2.2 - Proyecto estéreo

Para este caso, el resumen va a ser mucho más breve debido al poco interés que tiene el trabajar en formato estéreo para el presente proyecto. Únicamente se va a hacer mención a los aspectos de interés o que puedan tener que ver con ambos casos en común para poner en contexto el tipo de mezcla que se va a realizar y así poder poner en contexto ambas mezclas.

En la figura 5.7 se puede observar la disposición de los instrumentos a lo largo del espacio estéreo. Cada familia de instrumentos se ha colocado en un estrato vertical.

Como se aprecia, se ha buscado llenar el estéreo lo máximo posible. En primer lugar se ha colocado la batería como es usual, con el bombo al centro de la mezcla y los ambientes de batería cada uno en un extremo para formar el estéreo. El contrabajo se ha situado al centro de forma virtual, situando cada una de las dos pistas de las que está compuesto ligeramente paneada a cada lado de forma simétrica, creando la sensación de que éste se encuentra aproximadamente en el centro de la mezcla. Guitarra y piano se han colocado cada uno a un lado de la mezcla, buscando crear esa sensación de espacialidad al tener un timbre distinto en cada lado. Los vientos se han intercalado a lo lado del espacio, situando cada instrumento similar al lado opuesto de su igual. La voz se ha situado en el centro. Por último se ha situado los micrófonos de ambiente de sala repartidos por la mezcla.

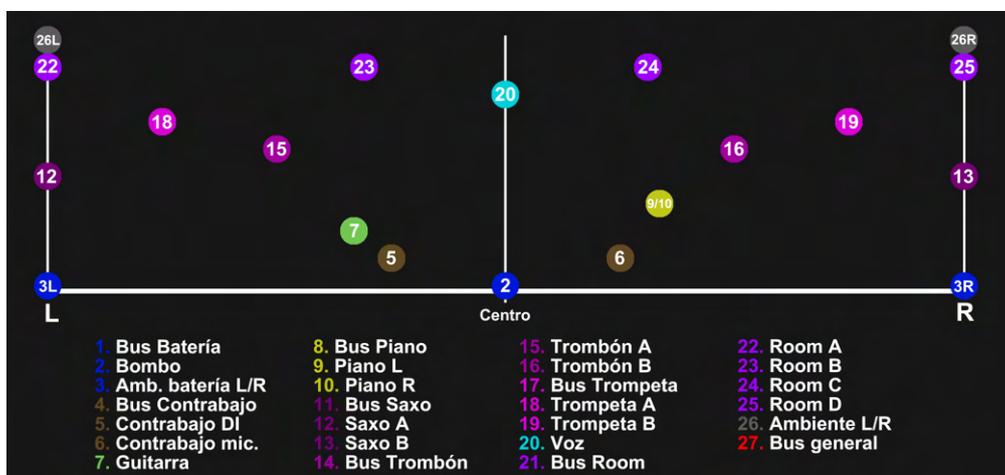


Figura 5.7: Diagrama con la ubicación espacial de cada instrumento y lista de canales a lo largo del espacio estéreo.

La agrupación de pistas y buses es la misma que en el caso práctico 1, así como los colores de cada pista y su representación en la figura 5.7.

En cada instrumento únicamente se ha alterado por canal el paneo y el volumen del mismo. Se ha decidido no aplicar ningún tipo de procesado por canal para intentar ajustar al máximo las similitudes en la mezcla con el caso práctico 2.1.

Por tanto, únicamente se ha insertado procesamiento en el bus general de mezcla. Para este bus, se ha seguido la misma cadena compuesta por: EQ → Compresor → Reverb → Limitador → Visualizador.

Cada procesador se ha utilizado con la misma finalidad el caso práctico 2.1, buscando realzar algo de brillo en la mezcla y controlar la zona de bajas frecuencias. El visualizador en este caso no es Ambisonics, y se ha utilizado para controlar tanto la imagen estéreo como la respuesta frecuencial y la sonoridad de la mezcla.

De esta forma, intentamos dejar las mezclas entre ambos casos lo más similares posibles, buscando marcar las diferencias entre ambas únicamente en el ámbito espacial.

El resultado se puede escuchar en la pista “Ambisonics - Ejemplo 2.2 (binaural)” de la carpeta “FINAL MIXES”, adjunta a este proyecto, en formato **.mp3 24 bits - 44.1 kHz**.

5.3 Caso práctico 3: Multipista sintético

Como último caso práctico se va a proceder a realizar una mezcla Ambisonics a partir de una producción propia hecha toda a través de *samplers* y síntesis. Por tanto, ninguno de los instrumentos que suenan ha sido grabado en la realidad. Para este ejemplo se trata de un proyecto de 6 pistas (5 mono y 1 estéreo), las cuáles se van a mover por el espacio sonoro de formas agresivas para buscar enfatizar esta característica espaciales que ofrece Ambisonics. La misma mezcla se hará en orden 3 y en orden 5, buscando evaluar las posibles diferencias entre ambos órdenes en el capítulo de Resultados, y la viabilidad de trabajar en órdenes superiores.

Ambas mezclas son totalmente idénticas salvo por la excepción de que en una se ha trabajado íntegramente en orden 3, y en la otra en orden 5.

En la figura 5.8 se puede observar la ventana sesión con las 6 pistas y sus envolventes de automatización. A continuación se enumeran los colores de las envolventes y el parámetro al que corresponden:

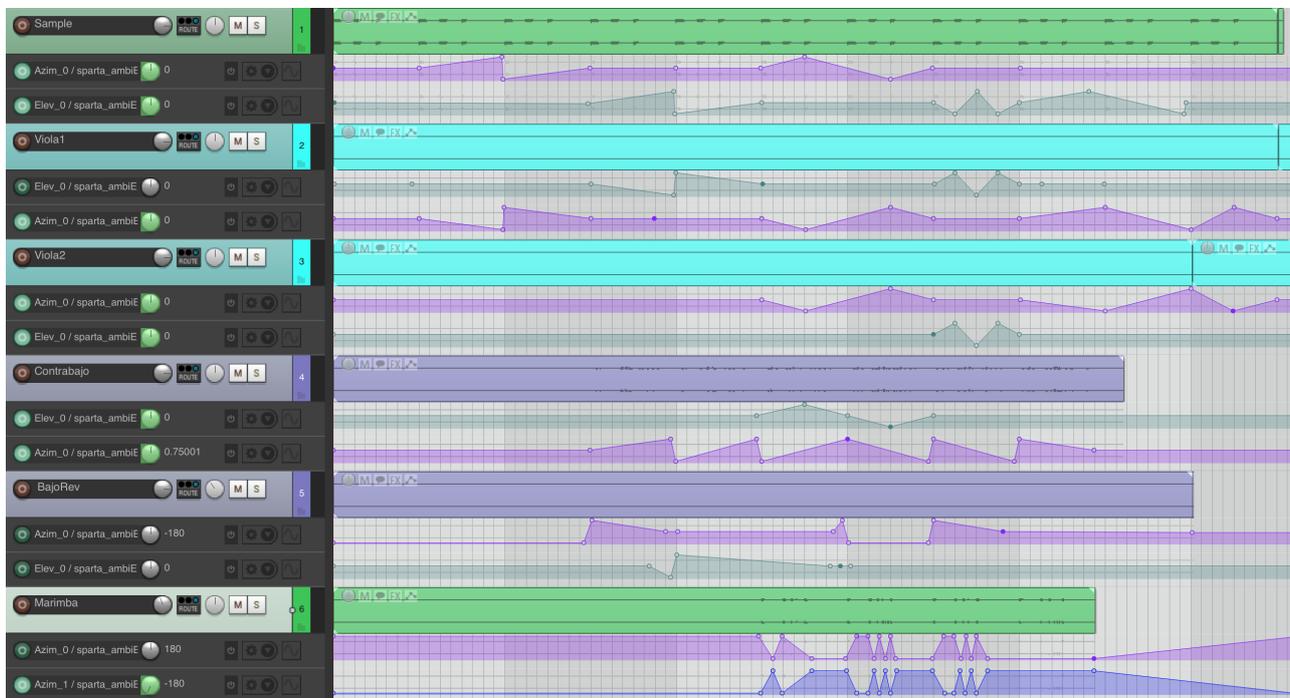


Figura 5.8: Ventana sesión del caso práctico 3 con todas las envolventes y su automatización.

- **Violeta.-** Las envolventes de color violeta o añil corresponden al parámetro *Azimuth* del canal

1. Esto quiere decir que para pistas monofónicas, controlara el movimiento de la fuente correspondiente al canal 1 a lo largo de los 360° en el plano horizontal.

- **Verde.-** Las envolventes de color verde corresponden al parámetro *Elevación* del canal 1. Esto quiere decir que para pistas monofónicas, controlara el movimiento de la fuente correspondiente al canal 1 a lo largo de los 360° en el plano vertical.
- **Azul.-** Las envolventes de color azul corresponden al parámetro *Azimuth* del canal 2. Esto quiere decir que para pistas estereofónicas, controlara el movimiento de la fuente correspondiente al canal 2 a lo largo de los 360° en el plano horizontal.

Todas las pistas son monofónicas excepto la última pista, Marimba, que se trata de una pista estéreo. Para diferenciarla del resto y por su poca presencia en la composición, únicamente se han movido ambas fuentes a lo largo del plano horizontal. Es por ello que para esta pista aparece la envolvente de color azul, correspondiente al parámetro *azimuth* del canal 2. El resto de pistas únicamente poseen un canal (mono).

Es importante ajustar el modo de envolvente de cada pista en modo *read* para que *Reaper* lea estas envolventes, editadas a ratón, sin modificarlas. También es importante tener activada la casilla de envolvente global, en la parte superior de *Reaper*, junto a la barra de transporte.

El ruteo de todas las pistas es similar a los casos prácticos anteriores. Todas las pistas son enviadas a un bus general llamado AMBISONICS, donde se realzan las frecuencias agudas perdidas durante el procesado, y se muestra un visualizador de energía Ambisonics. Tras esto, los 16/36 (orden 3 y orden 5) canales de la mezcla general son enviados a un bus de monitorización donde se colocará el decodificador Ambisonics a binaural con el orden correspondiente y se limitará la salida para evitar saturaciones.

Los movimientos de las distintas fuentes por el campo sonoro se han realizado buscando enfatizar los diferentes movimientos y buscando situar los distintos instrumentos en diferentes espacios de manera simultánea para poder apreciarlos por separado e intentar identificar el movimiento que sigue cada uno, siendo éste más sencillo en unos casos que en otros.

Las únicas diferencias apreciables a la hora de trabajar en orden 3 u orden 5 (o cualquier orden en realidad) es dimensionar las pistas al número de canales demandado por el orden de trabajo. En nuestro caso particular, las pistas y los buses para orden 3 son de 16 canales, mientras que para orden 5 el número asciende a 36 canales. La otra diferencia significativa será variar el orden de trabajo en todos los *plugins* Ambisonics, en nuestro caso serán el SPARTA AmbiENC disponible en cada canal para situar cada instrumento en el espacio sonoro, el IEM MultiEQ para la ecualización del bus general, el visualizador SPARTA SLDoA y el decodificador binaural SPARTA AmbiBIN.

El resultado se puede escuchar en las pistas “Ambisonics o3 - Ejemplo 3 (binaural)” y “Ambisonics o5 - Ejemplo 3 (binaural)” de la carpeta “FINAL MIXES”, adjunta a este proyecto, en formato **.mp3 24 bits - 44.1 kHz**.

Capítulo 6

Resultados

En el presente capítulo se van a evaluar de una forma subjetiva (y objetiva si se pudiese) los tres casos prácticos desarrollados en el capítulo 5. Resaltar el hecho de que se trata de material sonoro, donde la percepción depende especialmente de la subjetividad de quien lo escucha. Por tanto, la evaluación de resultados se va a realizar a través de un lenguaje que en muchos casos se basará en sensaciones o percepciones subjetivas que para otro sujeto pueden ser diferentes.

Para la escucha es imprescindible utilizar auriculares ya que se trata de material espacial decodificado para escucha binaural. Por otro lado, se recomienda el uso de auriculares de calidad, especialmente de diadema. El uso de HRTF personalizada o archivo de calibración de los auriculares mejorará notablemente la escucha, si bien en muchos casos es difícil disponer de ambos archivos. Para la evaluación de resultados se han utilizado los auriculares *Sennheiser*® *HD-25 Plus*, cuya hoja de especificaciones se expone a continuación:

Parámetro	Valor
Transductor	Dinámico
Tipo de auricular	Cerrado / On-ear
Impedancia	70Ohm
SPL máximo	120dB (1kHz, 1Vrms)
Respuesta en frecuencia	16 - 22000Hz
THD(1kHz)	<0,3 %
Capacidad de carga	200mW

6.1 Resultados caso práctico 1: Ambisonics + *spot*

El caso práctico 1 está basado en una mezcla en Ambisonics a partir de un material original grabado de forma híbrida en formato Ambisonics de orden 1, con dos *Sennheiser*® *AMBE0* para la posición del director/principal y otra en la zona de audiencia con el mismo micrófono, así como 4 micrófonos monofónicos para cada instrumento y dos más en formato L/R para el ambiente de la sala.

Para la ubicación de los instrumentos, y al partir de un campo sonoro definido gracias a los micrófonos Ambisonics, la ubicación de los instrumentos mediante los *panners* ha sido la de su posición original (en la medida de lo posible al carecer de medidas de campo). Esta posición por tanto no es muy agresiva y salvo por las reflexiones naturales de la sala, las cuáles son capturadas según su dirección gracias a los micrófonos Ambisonics, dichas posiciones se podrían haber realizado a través de un *panner*

estéreo y por tanto se podría haber replicado el resultado.

A pesar de esto, y cuando escuchamos el resultado de la mezcla, podemos apreciar un campo sonoro mucho más abierto, creando la sensación de que el oyente realmente se encuentra en la sala frente a los intérpretes. Esta sensación se ve acentuada por el sonido que llega desde la parte trasera desde el micrófono de audiencia y las reflexiones naturales de la sala. El resultado por tanto se traduce en una pieza con una imagen espacial muy rica, absorbiendo al oyente al centro de la misma y envolviéndolo desde todas las direcciones de la escucha.

Además, y gracias a los micrófonos *spot*, en ningún momento se pierden los matices propios de cada instrumento o se crea un sonido difuso y lejano. Al complementar la riqueza espacial de la captación Ambisonics con la precisión y definición de los micrófonos monofónicos, se consigue una pieza que alberga lo mejor de ambos mundos, con una definición de los instrumentos con gran presencia y una riqueza espacial espectacular.

Por tanto, y en conclusión para el análisis de esta pieza, podemos definir la técnica de grabación híbrida entre micrófonos estándar y microfónica Ambisonics, sumados a una mezcla Ambisonics con decodificación binaural, como un gran método de trabajo para grabaciones de música, especialmente para piezas acústicas donde la importancia de la acústica de la sala y el timbre de los instrumentos se hace tan necesaria.

Además, para concluir, se ha realizado una medición de la fase a lo largo de toda la canción y toda ella se encuentra en fase sin sufrir cancelaciones por la configuración multi micrófono.

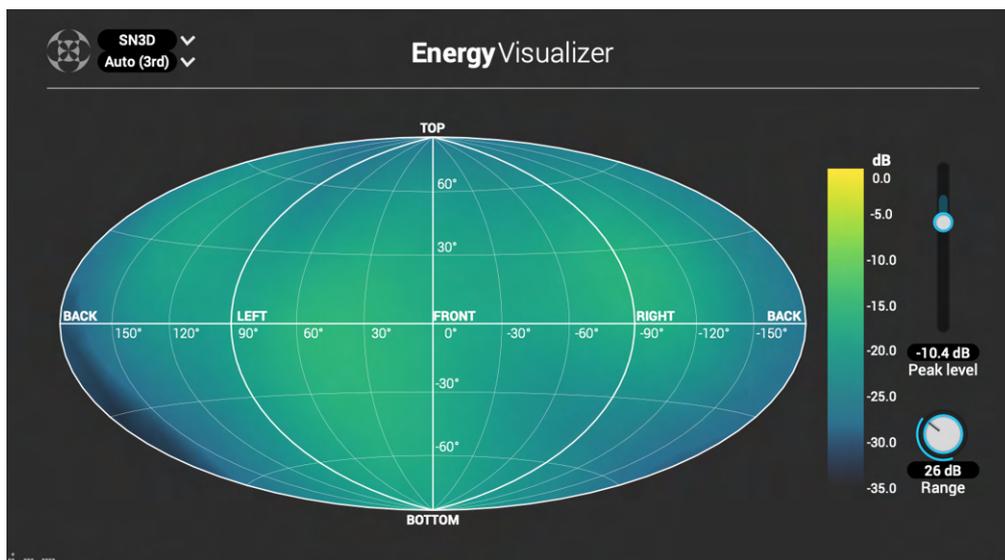


Figura 6.1: *IEM EnergyVisualizer* durante una parte de la mezcla del caso práctico 1.

6.2 Resultados caso práctico 2: Multipista acústico

Para este caso se ha partido de una sesión de grabación multi pista con microfónica monofónica estándar. A partir de ahí se han realizado dos mezclas, una en formato Ambisonics de orden 3 y otra en formato estéreo. Por tanto, para su análisis se va a proceder a evaluar cada una por separado en primer lugar, para posteriormente comentar una posible comparativa con diferencias y apreciaciones comunes.

6.2.1 Resultados Caso práctico 2.1 - Proyecto Ambisonics

Para la mezcla Ambisonics, y tal y como se ha mencionado en la sección anterior, únicamente se ha buscado situar los instrumentos a lo largo del campo sonoro simulando la ubicación normal de los músicos dentro de una *big band*. Escuchando la mezcla se puede apreciar una separación de los mismos muy notable dentro del campo sonoro. Prestando atención, uno es capaz de distinguir desde que dirección llega cada instrumento (o familia de instrumentos). El oyente escucha la pieza como si estuviese situado en la posición del director de la banda, con instrumentos colocados a su alrededor en un círculo casi completo. Además, varios de ellos se han situado con una altura positiva por encima del plano de escucha, simulando que están en algún tipo de grada donde se suele situar a los músicos. Esto consigue que cada instrumento esté más separado del resto, poseyendo cada uno de ellos una ubicación propia en el espacio, lo que facilita su ubicación dentro del mismo al no tener que compartir dicho espacio con otros instrumentos.

Por otro lado, y al trabajar con algún tipo de micrófono en configuración estéreo, es posible que si la ubicación real de los instrumentos no coincide con la ubicación virtual donde lo hemos situado en el campo sonoro, esto provoque conflictos en la dirección de llegada de dicho instrumento. En otras palabras, a pesar de haber situado un instrumento de manera virtual en una posición concreta, esta posición se puede volver difusa en la escucha debido a los micrófonos estéreo de la grabación. Para evitar este conflicto, se barajó la idea de eliminar estos micrófonos estéreo de la mezcla y trabajar únicamente con micrófonos monofónicos. Se descartó esta posibilidad por buscar que aún con dichos micrófonos estéreo, la mezcla Ambisonics alcanzase unos resultados satisfactorios, buscando así demostrar que es una buena técnica de mezcla aún con grabaciones que puedan no estar pensadas originalmente para su mezcla en Ambisonics.

El resultado es muy satisfactorio, sintiendo cada instrumento en su posición y pudiendo identificarlo con bastante precisión dentro del campo sonoro. Además, al tratarse de grabaciones con micrófonos monofónicos situados cerca del instrumento, éste se escucha con mucha “pegada” y definición, pudiendo situar cada instrumento al frente de la mezcla o pudiendo alejarlo con efectos como la reverberación o la cantidad de volumen. Esto nos ofrece la posibilidad de, no sólo situar cada instrumento en cualquier dirección de escucha, si no poder crear planos sonoros dentro de cada una de esas direcciones.

En definitiva el resultado es, cómo ya se ha dicho, satisfactorio, con una mezcla rica en matices tanto tímbricos como espaciales, además de muy homogénea en su conjunto y con la sensación de que toda la banda está sonando frente al oyente.

6.2.2 Resultados Caso práctico 2.2 - Proyecto estéreo

Para la mezcla estéreo, se siguió un procedimiento parecido a la mezcla Ambisonics, buscando que cada instrumento o familia de instrumentos tuvieran su propio espacio en la mezcla a fin de conseguir que éstos se identificasen claramente y no tuvieran que pelear con otros instrumentos que dificultasen su definición y les quitaran protagonismo en la mezcla. Aún así, esta tarea se vuelve mucho más complicada al tener un gran número de pistas y un espacio donde colocarlas mucho más limitado. Para Ambisonics tenemos un campo sonoro completo de 360 grados, o lo que es lo mismo, podemos situar la fuente sonora donde nosotros deseamos a lo largo de una esfera completa alrededor del oyente. Para una mezcla estéreo esto se limita a un espacio situado en el frontal del oyente de unos 180 grados y en un sólo plano horizontal. Es evidente que éste espacio es mucho más limitado y por ello una mezcla cómo esta con un gran número de pistas se hace mucho más difícil poder garantizar un espacio propio a cada uno de los instrumentos.

El resultado final, a pesar de haber intentado repartir los instrumentos a lo largo de todo el estéreo y

evitando en la medida de lo posible que éstos compartiesen el mismo espacio, es una mezcla que si bien mantiene la definición propia de cada instrumento, ésta se ve enmascarada por tener que compartir espectro frecuencial y espacio con otros instrumentos dentro de la mezcla. La apertura espacial de la mezcla es más limitada y la sensación por parte del oyente es que todo queda más difuso en la zona frontal de escucha, tal y como se podría esperar de una mezcla estéreo.

6.2.3 Resultados generales Caso práctico 2

Como vemos de manera gráfica en la figura 6.2, para la mezcla estéreo final codificada en tercer orden para la representación gráfica de su campo sonoro, y en la figura 6.3 para la mezcla Ambisonics, los campos sonoros en un momento puntual de ambas mezclas son muy diferentes en apertura y espacialidad. Es importante destacar que la escala para la representación de la mezcla Ambisonics se ha modificado al poseer esta menos nivel intrínseco y por tanto quedar difusa su representación. A pesar de ello, y cómo es lógico para una mezcla estéreo, ésta queda limitada entre los espacios de izquierda y derecha del oyente en un arco de 180 grados. Esto limita que pueda existir información sonora en la parte trasera del oyente, así como en los planos superior e inferior, tal y como se aprecia en ambas representaciones gráficas.

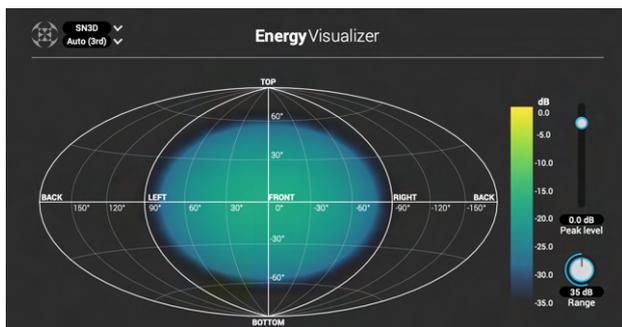


Figura 6.2: *IEM Energy Visualizer* para mezcla estéreo.

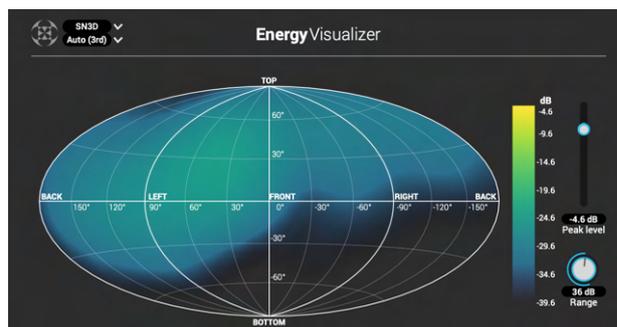


Figura 6.3: *IEM Energy Visualizer* para mezcla estéreo.

Desde un punto de vista sonoro, que al final es lo que importa en un material sonoro, se aprecian de manera rápida incluso para oídos menos experimentados una diferencia en la amplitud del campo sonoro. Si bien es cierto que la mezcla estéreo nos suena más cotidiana o familiar debido a que el material musical usual de consumo suele ser en formato estéreo, en el momento en el que uno escucha la mezcla Ambisonics en formato binaural, uno es capaz de reconocer enseguida sus virtudes espaciales. El oyente se siente envuelto de inmediato y empieza a poder desgranar cada uno de los instrumentos y su espacio dentro del campo sonoro.

Se trata de un método de mezcla y escucha al que es muy fácil acostumbrarse principalmente porque imita la manera natural en la que los humanos escuchamos en nuestro día a día. Enseguida entendemos si un sonido nos llega por una u otra dirección, y, si el usuario posee una HRTF propia y unos buenos auriculares calibrados, la precisión en la identificación de las distintas direcciones de llegada de los sonidos se vuelve superior, llegando a engañar al cerebro haciéndole creer que uno mismo se encuentra en la sala con los músicos durante la interpretación de la pieza.

6.3 Resultados caso práctico 3: Multipista sintético

Para el último caso práctico se ha realizado una misma mezcla en formato Ambisonics a partir de varias pistas sintetizadas y secuenciadas por ordenador y se ha buscado dotar a cada pista de movimientos agresivos y poco naturales a lo largo del campo sonoro, de forma que éstos movimientos fuesen identificables a pesar de no ser naturales. Esta mezcla, una vez hecha, se ha trabajado y exportado para Ambisonics de orden 3 y para orden 5 (ambos casos se han tratado de forma idéntica y únicamente se ha variado el orden Ambisonics de trabajo, tanto para los plugins de panning como a la hora de “mezclar” y decodificar a binaural), buscando determinar si existen diferencias evidentes entre ambos órdenes. Es importante destacar que para trabajar en orden 3 debemos utilizar buses de 16 canales, mientras que para orden 5 serán de 36 canales. Esto es importante porque también deberá tenerse en cuenta a la hora de la valoración por consumo de recursos en el sistema.

Para realizar una comparación lo más sencilla posible para el usuario, se ha realizado un vídeo denominado **Ambisonics - Ejemplo 3.mp4**, donde se han reproducido secuencialmente cada uno de los ejemplos con una representación gráfica del campo sonoro de cada uno, buscando tener tanto el audio como su representación gráfica de cara a la valoración (a pesar de lo limitado de este tipo de representaciones gráficas).

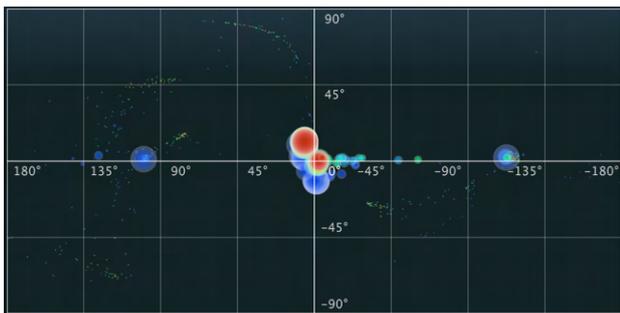


Figura 6.4: *IEM EnergyVisualizer* para orden 3.

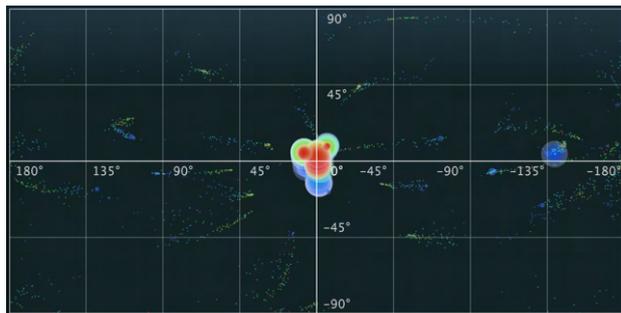


Figura 6.5: *IEM EnergyVisualizer* para orden 5.

Aunque se ha tratado de realizar la captura para la representación gráfica de cada uno de los dos órdenes, ésta no se ha podido realizar exactamente en el mismo instante, si bien la diferencia entre ambas es de milisegundos, se puede apreciar que no es exactamente el mismo. Aún con ello, en las figuras 6.4 y 6.5 tenemos un ejemplo de las diferencias entre ambos órdenes de trabajo, representado de forma gráfica. El ejemplo en orden 5 ofrece un mayor número de direcciones al estar formado por un mayor número de canales. Esto se aprecia en las zonas coloreadas en forma de nube de puntos en la figura 6.5. Estas zonas con contenido sonoro deben aparecer por errores en la codificación o por causas desconocidas, ya que en éste proyecto no se ha añadido reverberación a la mezcla, lo que podría justificar estas zonas por las reflexiones de la reverberación. Aún así, estas zonas ofrecen una representación de las distintas direcciones adicionales que se pueden reproducir en una mezcla de orden superior, teniendo zonas “ciegas” o direcciones no codificables en órdenes inferiores como el que se aprecia en la figura 6.4.

A nivel sonoro, las diferencias no son excesivamente diferenciables, si bien pueden llegar a apreciarse. Se supone que con unos auriculares con HRTF y de una gama de alta fidelidad con una reproducción de los detalles superior podrían llegar a apreciarse de una manera mucho más notable. Por otro lado el factor de la audición propia del oyente también influye en esta apreciación por lo que en todo momento se está hablando desde un punto subjetivo. Teniendo en cuenta estos detalles, y con nuestra escucha, se puede llegar a apreciar una mayor resolución espacial en el orden superior. Esto es apreciable especialmente en los desplazamientos laterales, donde se puede escuchar una mayor continuidad o fluidez

en el movimiento, al contrario que para orden 3 donde en estos desplazamientos se llega a apreciar una “discretización” de estos movimientos, como si éstos fueran a pequeños saltos. El hecho de que sea en los desplazamientos laterales donde se pueden llegar a apreciar diferencias es debido a que la escucha humana es mucho más precisa en la zona frontal y lateral de nuestra escucha. Además, y tal y como se comentó en la sección 2.1.2, el efecto del cono de difusión dificulta la identificación de sonidos en planos superiores e inferiores. Es una diferencia muy sutil pero reseñable en un análisis de este tipo. Por supuesto y como ya se ha dicho, estas sutiles diferencias pueden llegar a no ser apreciables por todo el mundo o con cualquier auricular.

Como contraposición al uso de un orden superior para nuestra mezcla Ambisonics, tenemos el consumo de recursos. Se han tomado 10 muestras aleatorias del porcentaje de consumo de la CPU para cada orden a lo largo de toda la canción, alojando unos resultados del 6.69% para orden 3, y del 10,14% para orden 5. Por supuesto es importante destacar que en estos ejemplos el uso de *plugins* de procesado, exceptuando los Ambisonics, ha sido nulo, cosa que en una sesión real harían falta muchos de estos para conseguir una mezcla con una calidad aceptable. Esto hace que determinar el orden Ambisonics de trabajo venga limitado por los recursos de nuestro ordenador. Para grandes proyectos con alto número de pistas, o donde se trabaje con procesadores con alto consumo de recursos, como el *IEM RoomEncoder* por ejemplo, el uso de órdenes superiores puede hacer que el *playback* de nuestra sesión se vea ralentizado y la reproducción se dé a saltos, algo nada deseable. Por ello, y como situación de compromiso, se puede concluir que trabajar en orden 3 puede ser un buen punto de partida, donde la resolución espacial es muy alta, llegando incluso a ser inapreciable la diferencia con órdenes superiores. Por supuesto, no se recomienda trabajar en órdenes menores a 3, donde la resolución espacial de éstos se ve muy limitada y perderemos gran parte del potencial que nos ofrece trabajar en formato Ambisonics.

Capítulo 7

Conclusiones y trabajos futuros

7.1 Conclusiones

A partir de los objetivos planteados se pueden sacar varias conclusiones sobre el potencial de la tecnología Ambisonics.

En primer lugar se han visto y comprendido varios conceptos esenciales tales como: qué es Ambisonics, cómo funciona esta tecnología y sobre qué fundamentos teóricos y físicos se sustenta. En definitiva se ha comprendido todo lo relacionado sobre Ambisonics desde un punto de vista conceptual y de principios de funcionamiento.

En segundo lugar se han planteado y diseñado de manera satisfactoria distintos sistemas y aplicaciones para el uso de Ambisonics. Dichos sistemas han sido una virtualización de un sistema de audio para cine 5.1 en formato Ambisonics, donde los resultados son notablemente satisfactorios, mejorando la escucha de cine a través de auriculares notablemente y dotando a esta de una espacialidad mucho mayor. Además, y como complemento a la escucha binaural, se ha implementado un sistema de seguimiento *head-tracker* basado en Arduino y utilizable desde nuestro DAW para escucha Ambisonics con auriculares. Por otro lado se han diseñado e implementado una serie de sistemas de reproducción 3D para Ambisonics a partir del laboratorio GTAC de la UPV. Estos sistemas han ofrecido unos resultados en la escucha dispares, necesitando una optimización y ajuste de los mismos para una correcta reproducción de todo el material Ambisonics.

En tercer lugar se han repasado los diferentes sistemas de grabación de campo en formato Ambisonics, especialmente focalizando el proyecto en comprender y entender cómo grabar Ambisonics con los micrófonos disponibles en el laboratorio de la UPV, si bien estos conceptos se pueden extrapolar a cualquier micrófono de esta tecnología.

Por último se han realizado una serie de mezclas independientes en formato Ambisonics partiendo de diferentes materiales de inicio, tales como grabaciones Ambisonics, monofónicas o estéreo, y se ha visto cómo trabajar a lo largo de todo el proceso de mezcla en formato Ambisonics, factores determinantes a tener en cuenta o tecnologías y técnicas disponibles. Además, se han evaluado brevemente y de una forma subjetiva los resultados de dichas mezclas.

Por todo lo mencionado se puede concluir que Ambisonics se postula como una de las principales y más interesantes tecnologías de audio espacial actuales, gracias a su modularidad, personalización e independencia entre las fases de grabación, edición y reproducción, lo que la hace ideal para su integración dentro de sistemas como consolas de videojuegos.

7.2 Propuestas de trabajo futuras

Como propuestas de trabajo futuras se pueden ver diferentes ramas de trabajo a partir del presente proyecto, entre las que destacan:

- Ampliación del presente proyecto mediante la realización de diferentes grabaciones de campo en formato Ambisonics y la comparación de estas grabaciones con otros sistemas de grabación espacial.
- Mejora en la codificación de formato A-B a través de la medición personalizada de cada micrófono y generación de su archivo de corrección personalizado.
- Realización de un sistema de virtualización acústica de recintos a partir de Ambisonics y sistemas de convolución a partir de la IR del recinto.
- Diseño y desarrollo de un sistema de reproducción optimizado y calibrado de tercer orden dentro de las instalaciones de la Universitat Politècnica de València para reproducción de material Ambisonics.

Bibliografía

- [1] Ambisonic, *Wikipedia*, 2013.
Disponible en <https://en.wikipedia.org/wiki/Ambisonic>
- [2] ¿Cómo ubicamos los sonidos en el espacio? Binauralidad y Teoría Dúplex, *Hispasonic*, 2017.
Disponible en <https://bit.ly/2LSw5H4>
- [3] Sensitivity of Human Ears, *Hyperphysics*.
Disponible en <http://hyperphysics.phy-astr.gsu.edu/hbasees/Sound/earsens.html>
- [4] CETTA, P., *Localización espacial del sonido*, 2017. Universidad Centroamericana UCA (Managua, Nicaragua).
Disponible en <http://www.ub.edu/pa1/node/66>
- [5] ¿Cómo ubicamos los sonidos en el espacio? HRTF y otros indicios, *Hispasonic*, 2017.
Disponible en <https://bit.ly/2YP71HN>
- [6] Fonógrafo, *Wikipedia*, 2019.
Disponible en <https://es.wikipedia.org/wiki/Fon%C3%B3grafo>
- [7] Cilindro de fonógrafo, *Wikipedia*, 2019.
Disponible en https://es.wikipedia.org/wiki/Cilindro_de_fon%C3%B3grafo
- [8] Gramófono, *Wikipedia*, 2019.
Disponible en <https://es.wikipedia.org/wiki/Gram%C3%B3fono>
- [9] Disco de vinilo, *Wikipedia*, 2019.
Disponible en https://es.wikipedia.org/wiki/Disco_de_vinilo
- [10] Théâtrophone, *Wikipedia*, 2019.
Disponible en <https://en.wikipedia.org/wiki/Th%C3%A9%C3%A2trophone>
- [11] Principles of the Blumlein stereo technique, *DPA*, 2016.
Disponible en <https://bit.ly/2TUiOAM>
- [12] Essence Reveals The History of Stereophonic Sound, *Essence*, 2014.
Disponible en <https://bit.ly/2KiAPsZ>
- [13] Sistema de sonido estereofónico multicanal con y sin acompañamiento de imagen , *UIT*, 1994.
Disponible en https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.775-1-199407-S!PDF-S.pdf
- [14] NVSonic 3DOF Head Tracker, *GitHub*, 2019.
Disponible en <https://github.com/trsonic/nvsonic-head-tracker>

- [15] ZALLES, G., *NVSonic Headtracker NYU*, 2018. NYU Immersive Audio Group, 2018.
Disponibile en <https://wp.nyu.edu/immersiveaudiogroup/2018/01/26/nvsonic-headtracker-nyu/>
- [16] Dolby Atmos: Why it's cool, how it works and how to get it, *Cnet*, 2017.
Disponibile en <https://www.cnet.com/news/dolby-atmos-why-its-cool-how-it-works-and-how-to-get-it/>
- [17] Real/Complex Spherical Harmonic Transform, Gaunt Coefficients and Rotations, *Archontis Politis*, 2016.
Disponibile en <https://www.mathworks.com/matlabcentral/fileexchange/43856-real-complex-spherical-harmonic-transform-gaunt-coefficients-and-rotations>
- [18] A format to B format conversion, *Angelo Farina*, 2015.
Disponibile en <http://pcfarina.eng.unipr.it/Public/B-format/A2B-conversion/A2B.htm>
- [19] Explicit formulas for High Order Ambisonic, *Angelo Farina*, 2017.
Disponibile en pcfarina.eng.unipr.it/Aurora/HOA_explicit_formulas.htm
- [20] Audio Monitor VST, *Google*.
Disponibile en <https://resonance-audio.github.io/resonance-audio/>
- [21] MØLLER, H., *Fundamentals of Binaural Technology*, 1992. Aalborg University (Aalborg, Denmark).
Disponibile en https://www.researchgate.net/publication/223265318_Fundamentals_of_binaural_technology
- [22] Ambisonic reproduction systems, *Wikipedia*, 2019.
Disponibile en https://en.wikipedia.org/wiki/Ambisonic_reproduction_systems
- [23] SOLVANG, A., *Spectral Impairment for Two-Dimensional HigherOrder Ambisonics*, 2008. Norwegian University of Science and Technology (Trondheim, Norway).
Disponibile en https://www.researchgate.net/publication/223265318_Fundamentals_of_binaural_technology
- [24] ZOTTER, F., FRANK, M., *AMBISONIC, A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*, 2019. Institute of Electronic Music and Acoustics University of Music and Performing Arts (Graz, Austria) .
- [25] ZOTTER, F., FRANK, M., *All-Round Ambisonic Panning and Decoding*, 2012. Institute of Electronic Music and Acoustics University of Music and Performing Arts (Graz, Austria) .
- [26] BATKE, J. M., KEILER, F., *Using vbat-derived panning functions for 3d ambisonics decoding*, 2010. 2nd International Symposium on Ambisonics and Spherical Acoustics(Paris, Francia) .
- [27] PULKKI, V., *Virtual sound source positioning using vector base amplitude panning*, 1997. Laboratory of Acoustics and Audio Signal Processing, HUT (Helsinki, Finlandia)
- [28] Ambisonic, Channel Formats, *Ambisonic.info*, 2011.
Disponibile en <https://ambisonic.info/ambisonics/channels.html>
- [29] BARTON, G., *Ambisonics: An Incomplete Glossary*, 1982. MCS Review Online Reprints.
Disponibile en <http://decoy.iki.fi/dsound/ambisonic/motherlode/reprint007.html>
- [30] DOBSON, R., *The AMB Ambisonic File Format*, 2012.
Disponibile en <http://rwdobson.com/bformat.html>

- [31] NACHBAR, C., ZOTTER, F., DELEFLIE, E., SONTACCHI, A., *AMBIX - A suggested Ambisonic format*, 2011. Ambisonic Symposium 2011(Lexington, KY)
- [32] Apple Core Audio Format Specification 1.0, *Apple Developers*, 2011.
Disponible en <https://apple.co/3t6pEFU>
- [33] Wave Field Synthesis, *Wikipedia*, 2013.
Disponible en https://en.wikipedia.org/wiki/Wave_field_synthesis
- [34] Spatially Oriented Format for Acoustics, *Sofa Conventions 0.5*.
Disponible en <https://sourceforge.net/projects/sofacoustics/>
- [35] Dolby Setup Guides, *Dolby*.
Disponible en <https://www.dolby.com/us/en/speaker-setup-guides/index.html>
- [36] ALLRADecoder Guide, *IEM plugin suite*.
Disponible en <https://plugins.iem.at/docs/allradecoder/>
- [37] Performing not linear processing (De-noising, Compression, Limiter, etc.) on High Order Ambisonics signals using Adobe Audition CC and the SPS approach, *Angelo Farina*, 2020.
Disponible en <http://pcfarina.eng.unipr.it/Aurora/Ambisonics-Denoising.htm>
- [38] Introducing SPS format and its first practical implementation, called Mach1, *Angelo Farina*, 2018.
Disponible en <http://pcfarina.eng.unipr.it/SPS-conversion.htm>
- [39] Explorers of Ambisonics - Paul Col.
Disponible en <https://www.creativefieldrecording.com/2017/03/21/explorers-of-ambisonics-paul-col/>
- [40] *Rode NT-SF1*.
Disponible en <https://www.soundonsound.com/reviews/rode-nt-sf1>
- [41] *Harpex v1.6 - Manual*.
Disponible en <https://harpex.net/manual.pdf>
- [42] *em32setGain - User Manual*.
Disponible en <https://bit.ly/35nH6uN>
- [43] *EigenUnits® - User Manual*.
Disponible en <https://bit.ly/3dPwFnv>
- [44] *EigenStudio® - User Manual*.
Disponible en <https://bit.ly/34kr18o>
- [45] Learn Ambisonics with John Escobar - *John Escobar*
Disponible en <https://bit.ly/39rwbDE>
- [46] Abletones Big Band Multitrack Files - *Telefunken Elektroakustik*
Disponible en <https://bit.ly/3puSSMD>

Anexos

Anexo A

Fórmulas para armónicos de orden superior (HOA).[19]

ACN	Orden	Nombre	Representación Azimuth/Elevation
0	0	W	1
1	1	Y	$\sin(a) * \cos(e)$
2	1	Z	$\sin(e)$
3	1	X	$\cos(a) * \cos(e)$
4	2	V	$\sqrt{3/4} * \sin(2a) * (\cos(e))^2$
5	2	T	$\sqrt{3/4} * \sin(a) * \sin(2e)$
6	2	R	$(1/2) * (3(\sin(e))^2 - 1)$
7	2	S	$\sqrt{3/4} * \cos(a) * \sin(2e)$
8	2	U	$\sqrt{3/4} * \cos(2a) * (\cos(e))^2$
9	3	Q	$\sqrt{5/8} * \sin(3a) * (\cos(e))^3$
10	3	O	$\sqrt{15/4} * \sin(2a) * \sin(e) * (\cos(e))^2$
11	3	M	$\sqrt{3/8} * \sin(a) * \cos(e) * (5(\sin(e))^2 - 1)$
12	3	K	$(1/2) * \sin(e) * (5(\sin(e))^2 - 3)$
13	3	L	$\sqrt{3/8} * \cos(a) * \cos(e) * (5(\sin(e))^2 - 1)$
14	3	N	$\sqrt{15/4} * \cos(2a) * \sin(e) * (\cos(e))^2$
15	3	P	$\sqrt{5/8} * \cos(3a) * (\cos(e))^3$
16	4		$\sqrt{35/64} * \sin(4a) * (\cos(e))^4$
17	4		$\sqrt{35/8} * \sin(3a) * \sin(e) * (\cos(e))^3$
18	4		$\sqrt{5/16} * \sin(2a) * (\cos(e))^2 * (7(\sin(e))^2 - 1)$
19	4		$\sqrt{5/32} * \sin(a) * \sin(2e) * (7(\sin(e))^2 - 3)$
20	4		$1/8 * (35\sin(e)^4 - 30\sin(e)^2 + 3)$
21	4		$\sqrt{5/32} * \cos(a) * \sin(2e) * (7(\sin(e))^2 - 3)$
22	4		$\sqrt{5/16} * \cos(2a) * (\cos(e))^2 * (7(\sin(e))^2 - 1)$
23	4		$\sqrt{35/8} * \cos(3a) * \sin(e) * (\cos(e))^3$
24	4		$\sqrt{35/64} * \cos(4a) * (\cos(e))^4$

*En la figura 3.1 el número en color rojo en la parte superior derecha de cada armónico esférico corresponden al ACN.

Anexo B

Coordenadas esféricas del Laboratorio de Percepción Espacial del Sonido

Altavoz	Azimuth (°)	Radio (m)	Elevación (°)	Altavoz	Azimuth (°)	Radio (m)	Elevación (°)
1	0/360	2	0	45	140	2	0
2	355	2	0	46	135	2	0
3	350	2	0	47	130	2	0
4	345	2	0	48	125	2	0
5	340	2	0	49	120	2	0
6	335	2	0	50	115	2	0
7	330	2	0	51	110	2	0
8	325	2	0	52	105	2	0
9	320	2	0	53	100	2	0
10	315	2	0	54	95	2	0
11	310	2	0	55	90	2	0
12	305	2	0	56	85	2	0
13	300	2	0	57	80	2	0
14	295	2	0	58	75	2	0
15	290	2	0	59	70	2	0
16	285	2	0	60	65	2	0
17	280	2	0	61	60	2	0
18	275	2	0	62	55	2	0
19	270	2	0	63	50	2	0
20	265	2	0	64	45	2	0
21	260	2	0	65	40	2	0
22	255	2	0	66	35	2	0
23	250	2	0	67	30	2	0
24	245	2	0	68	25	2	0
25	240	2	0	69	20	2	0
26	235	2	0	70	15	2	0
27	230	2	0	71	10	2	0
28	225	2	0	72	5	2	0

Altavoz	Azimuth (°)	Radio (m)	Elevación (°)	Altavoz	Azimuth (°)	Radio (m)	Elevación (°)
29	220	2	0	73	0/360	1,55	-50
30	215	2	0	74	315	1,55	-50
31	210	2	0	75	270	1,55	-50
32	205	2	0	76	225	1,55	-50
33	200	2	0	77	180	1,55	-50
34	195	2	0	78	135	1,55	-50
35	190	2	0	79	90	1,55	-50
36	185	2	0	80	45	1,55	-50
37	180	2	0	81	0/360	1,3	44
38	175	2	0	82	315	1,3	44
39	170	2	0	83	270	1,3	44
40	165	2	0	84	225	1,3	44
41	160	2	0	85	180	1,3	44
42	155	2	0	86	135	1,3	44
43	150	2	0	87	90	1,3	44
44	145	2	0	88	45	1,3	44

Anexo C

Sistemas de reproducción Ambisonics en el Laboratorio

C.1 Sistemas Dolby Surround

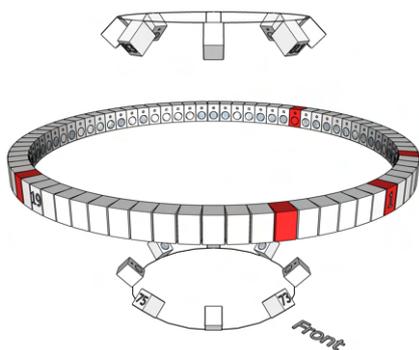


Figura C.1: Dolby 5.0.

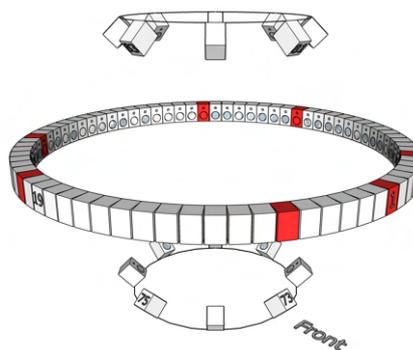


Figura C.2: Dolby 7.0.

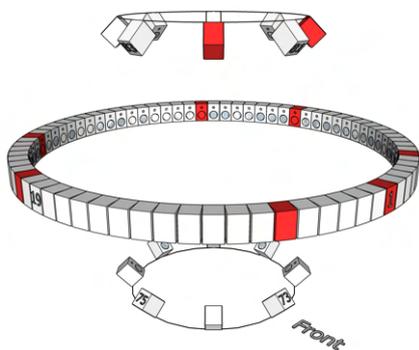


Figura C.3: Dolby 9.0.

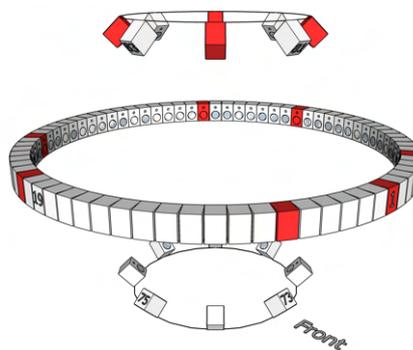


Figura C.4: Dolby 11.0.

C.2 Sistemas Ambisonics

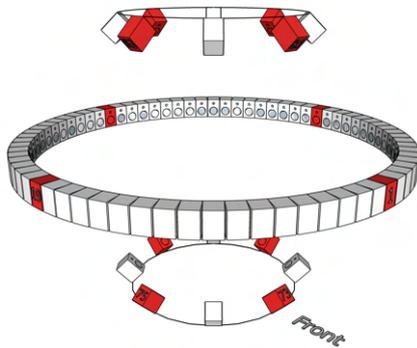


Figura C.5: 1^{er} orden (90°).

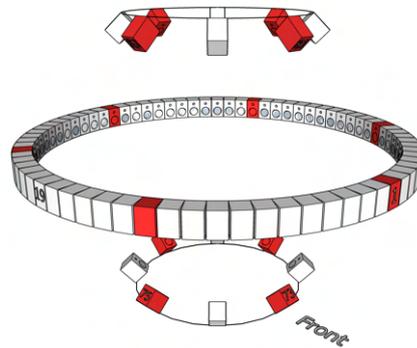


Figura C.6: 2^o orden (60°).

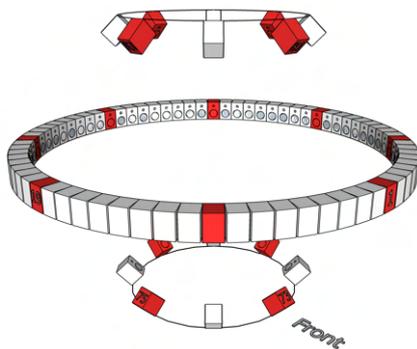


Figura C.7: 3^{er} orden (45°).

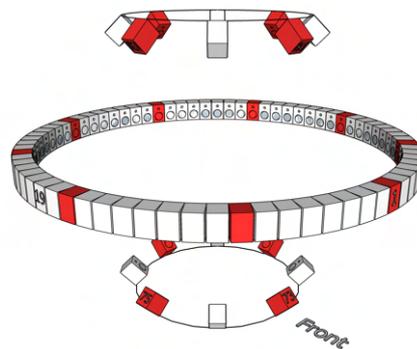


Figura C.8: 4^o orden (40°).

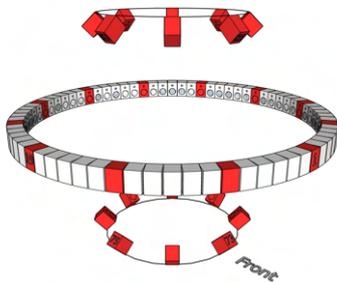


Figura C.9: 5^o orden (30°).

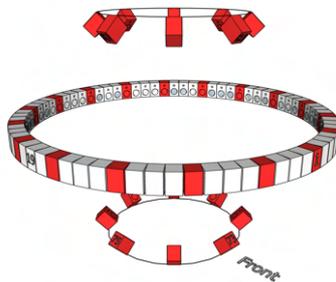


Figura C.10: 6^o orden (20°).

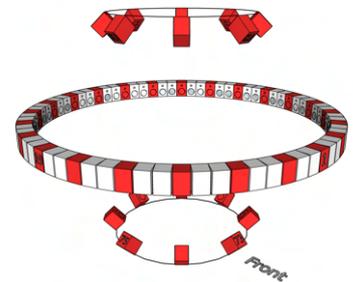


Figura C.11: 7^o orden (15°).

